COMPUTATIONAL ANALYSIS OF SIGNATURES OF SELECTION IN GENES FOR EGG PRODUCTION AND GROWTH IN POULTRY

EUNICE WAMBUI WAINAINA

MASTER OF SCIENCE

(Bioinformatics and Molecular Biology)

JOMO KENYATTA UNIVERSITY OF

AGRICULTURE AND TECHNOLOGY

2017

Computational Analysis of Signatures of Selection in Genes for Egg Production and Growth in Poultry

Eunice Wambui Wainaina

A thesis submitted in partial fulfillment for the Degree of Master of Science in Bioinformatics and Molecular Biology in the Jomo Kenyatta University of Agriculture and Technology

2017

DECLARATION

This thesis is my original work and has not been presented for degree in any other institution.

Signature... Date...

Eunice Wambui Wainaina

This thesis has been submitted for examination with our approval as university supervisors.

Signature... Date...

Dr. Sheila Cecily Ommeh, PhD,

JKUAT, Kenya

Signature... Date...

Prof. Daniel Kariuki, PhD,

JKUAT, Kenya

Signature... Date...

Dr. Jacqueline Lichoti, PhD,

DVS, Ministry of Agriculture, Livestock, and Fisheries, Kenya

DEDICATION

Special dedication to my parents Dorcas Wainaina and Stephen Wainaina. To my sister

Joan, brothers Charles, Philip, James, Joram, and to my niece Julie.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

3D	Three Dimensions
BLAST	Basic Local Alignment Search Tool
BLOSUM62	Blocks Substitution Matrix 62
bp	base pair
cAMP	Cyclic Adenosine Mono Phosphate
CRFs	Conditional Random Fields
Cu-Zn SoD	Copper-Zinc superoxide dismutase
DAG	Di-acyl glycerol
dN/dS	Ratio of rate of non-synonymous mutations to synonymous mutations
DNA	Deoxyribonucleic Acid
E Value	Expectation Value
Fe SoD	Iron Superoxide dismutase
GDP	Gross Domestic Product
GH	Growth hormone
GHR	Growth Hormone Receptor
IGF1	Insulin-like growth factor1
IGF1R	Insulin-like growth factor1 receptor
IRS1	Insulin Receptor Substrate 1
kDa	Kilo daltons
LD	Linkage disequilibrium
LRH	Long Range Haplotypes

LRT	Likelihood Ratio Test
MAFFT	Multiple Alignment using Fast Fourier Transform
MAPK	Mitogen-Activated Protein Kinase
MnSoD	Manganese Superoxide dismutase
mRNA	messenger Ribonucleic acid
mRNA	messenger Ribonucleic Acid
MSA	Multiple Sequence Alignment
MTT	Multiple Template Threading
MUSCLE	Multiple Sequence Comparison by Log Expectation
nr	non redundant
PAML	Phylogenetic Analysis using Maximum Likelihood
PDB	Protein Data Bank
PKC	Protein Kinase C
PRANK	Probabilistic Alignment Kit
PRL	Prolactin
SNP	Single Nucleotide Polymorphism
STAT	Signal Transducers and Activators of Transcription
VIP1	Vasoactive Intestinal Peptide 1
VIPR1	Vasoactive Intestinal Peptide Receptor 1

ABSTRACT

Poultry are highly variable phenotypically as a result of natural selection and domestication. This leaves behind signatures of selection which may be used for genetic improvement of poultry through selective breeding. The aim of this study was to perform analysis of signatures of selection at candidate genes for egg production and growth. Genes selected for egg production were prolactin, vasoactive intestinal peptide 1 and vasoactive intestinal peptide receptor 1. Genes selected for growth were growth hormone, growth hormone receptor, insulin-like growth factor 1, and insulin-like growth factor 1 receptor. A reciprocal BLASTp using BLOSUM62 substitution matrix was performed to identify the homologs. Orthologs with an expectation value greater than 1e-10 were selected for further analysis. Thereafter, multiple sequence alignment was performed using MUSCLE which is based on an iterative algorithm. Phylogeny construction was then done using Distance-based FastME followed by analysis using codon-based models in PAML. Likelihood ratio test was used to detect positive selection followed by posterior probability using Bayes Empirical Bayes Analysis to identify the sites under selection. Web-based servers: Raptor X and DeepAlign were used to predict 3D structures and compare the structures, respectively. This led to identification of purifying selection in all lineages in vasoactive intestinal peptide receptor 1, while in prolactin, there was positive selection in poultry. Additionally, purifying selection was detected in poultry lineages for Growth Hormone and Growth Hormone Receptor genes. Insulin-like growth factor 1 receptor (IGF1R) had positive selection on amino acid Isoleucine at position 460 on Receptor L domain.

The positive selection on IGF1R may be used as a molecular marker in improving growth of poultry through molecular breeding. The computational approach is fast and accurate and may be used as an additional tool in genetic improvement of poultry.

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

Chickens, quails and guinea fowls are poultry which are domesticated by humans for production of eggs, meat, or feathers. Poultry is classified in the Kingdom *Animalia*, Phylum *Chordata*, Subphylum *Vertebrata*, Class *Aves*, Order *Galliformes* (Al-Nasser *et al*., 2007). The chickens belong to the Family *Phasianidae*, subfamily *Phasianinae*, Genus *Gallus*. There are four species of the genus *Gallus* which include *Gallus gallus*, *Gallus varius*, *Gallus lafayettei* and *Gallus sonnerati* and three sub species of *Gallus gallus* which are *Gallus gallus gallus*, *Gallus gallus spadiceus* and *Gallus gallus bankira* (Al-Nasser *et al*., 2007).

Though the introduction of chicken domestication in Africa has not been documented well, there are postulations that chickens were introduced to Africa first via Egypt then they dispersed southwards through River Nile (Mwacharo *et al*., 2013).

In East Africa, there is archaeological evidence of chickens in the mid seventeeth century BC in Sudan (Houlihan & Goodman, 1986), 800 AD in coastal Kenya (Marshall, 2000) and in Akameru and Cyinkomane in Rwanda (MacDonald, 1992). Subsequent dispersion to other parts of Africa however remains unclear (Mwacharo *et al*., 2013).

On the other hand, the quail has its origin in Northern and southern Asia and its dispersal to other parts of the world is through its ability to fly in short distances (Hosomichi *et al*., 2006). There are two avian *Coturnix* species: *Coturnix japonica* (Japanese quail) and *coturnix coturnix* (common quail). Of the two quail species, the Japanese quail is the domesticated one, while the common quail remained a wild bird (Hosomichi *et al*., 2006). The Japanese quail first migrated to North America, Europe and Middle East in the 20th century (Hosomichi *et al*., 2006).

The guinea fowl is a native of Africa. There are footprints and a clutch of eggs that have a close comparison to those of the guinea fowl which were found in Laetoli in Tanzania (Leakey and Hay, 1979). Therefore, the guinea fowl is believed to have originated from Tanzania.

1.2 Chickens for egg and meat production

Chickens account for the largest livestock species reared by man probably because they have little to less cultural or religious restrictions (Tadelle *et al*., 2003). There are currently 32 million poultry in Kenya (Agricultural sector development strategy, 2010-2020). Indigenous chickens form 76%, 22% is comprised of layers, broilers, and the breeding stock. The rest of the domestic birds like quails, guinea fowls, ducks, geese, turkeys, and pigeons make up 2% of the poultry population (Olwande *et al*., 2010).

In Kenya 70% of the population resides in the rural areas with 90% engaging in poultry production (FAO, 2014). Majority of the poultry farmers are women and the youth (Kingori *et al*., 2010; FAO, 2014).

According to FAO (2010), indigenous chickens lay an average of 40-60 eggs per year. These eggs are laid in 3-4 clutches with each clutch consisting of 12-20 eggs. Each egg weighs 25-49 g with the exception of the egg laid by the naked neck which is 52g (Olwande *et al.*, 2010). Indigenous chickens are also slow in maturing consequently influencing meat production (FAO, 2010).

The low production of eggs and meat in indigenous chickens has been attributed to poor housing, diseases, poor nutrition, and lack of appropriate breeds (Okeno *et al.*, 2012).

Kenya produces about 1.3 billion poultry eggs and 20 tonnes of meat in a year (Agricultural sector development strategy, 2010). Despite the indigenous chicken having a low production, they contribute 47% and 55% to the Kenya's egg and meat production respectively (King'ori, 2010).

There has been an increase in the demand for poultry products owing to the increase in human population, higher incomes, and higher standards of living (FAO, 2002).

Understanding the various production systems is important in identifying the needs of farmers, marketers, and consumers. This is crucial in formulating a holistic breeding program that aims at improving indigenous poultry (Okeno *et al.*, 2012). The production is however not meeting the demand by the population (USAID, 2010).

However, indigenous poultry are highly variable phenotypically in size, earlobes, beaks, plumage color, feather morphology and pattern, skin color, comb type, live weights, and

egg production, among other traits (Kingori *et al*., 2010). This variation provides an opportunity for improvement through selective breeding.

The distinct ecotypes of indigenous chickens have been named according to their phenotypic descriptions and they include, naked neck, frizzled feathered, barred feathered, feathered shanks, bearded, dwarf sized, kuchi, bantam, and rumpless (Moraa *et al*., 2015).

Most indigenous poultry are reared under small extensive production system (Okeno *et al*., 2012). In an attempt to improve production of eggs and meat in this system, local farmers use various observable traits like selection of egg size while others use selection of chicken for their preferred traits based on the performance history to influence successive generations. In some cases, cross breeding of indigenous poultry with exotic breeds is done which has not been successful because although production is generally improved, there is loss of broodiness making the poultry incapable of reproducing naturally (Okeno *et al*., 2012) (FAO, 2014). This crossbreeding is also accompanied by dilution of the appearance of the indigenous poultry which leads to rejection of the products by consumers (FAO, 2014). The costs of production also go up as the resultant crossbreeds are not adapted to scavenging conditions and there are also higher mortalities. These challenges are attributed to inappropriate technologies resulting from lack of understanding of the characteristics of the different production systems and the lack of information on consumer preferences and market dynamics (Okeno *et al*., 2012). Indiscriminate cross breeding is also threatening the existence of indigenous poultry which may lead to their extinction (Besbes, 2009).

The table below shows the preferred chicken traits by farmers, marketers and consumers that may help breeders in improving indigenous chickens to meet the demands (Okeno *et al.*, 2012).

Table 1.1: Index ranking of traits perceived by farmers, marketers and consumers to be of economic importance.

| | Index ranking | | | | | | | | | | | | | | |
| | Farmers | | | | | Marketers | | | | | Consumers | | | | |
Traits	1	2	3	Sum	Index	1	2	3	Sum	Index	1	2	3	Sum	Index
Growth rate	12.64	7.03	6.86	26.53	0.512	-	-	-	-	-	-	-	-	-	-
Body size	5.51	9.06	11.52	26.10	0.503	8.33	15.68	27.02	51.03	0.152	10.38	18.24	20.44	49.06	0.154
Egg number	7.63	7.96	10.70	26.29	0.507	-	-	-	-	-	-	-	-	-	-
Fertility	15.92	6.10	3.68	25.70	0.500	-	-	-	-	-	-	-	-	-	-
Prolificacy	5.73	13.02	-	19.08	0.368										
Disease resistance	14.87	6.973	4.25	26.08	0.503	-	-	-	-	-	-	-	-	-	-
Heat tolerance	3.17	5.17	10.62	18.97	0.366	-	-	-	-	-	-	-	-	-	-
Drought tolerance	1.17	4.13	4.90	10.20	0.200	-	-	-	-	-	-	-	-	-	-
Broodiness	3.40	9.18	6.21	18.79	0.362	-	-	-	-	-	-	-	-	-	-
Mothering ability	4.51	6.10	7.92	18.54	0.357	-	-	-	-	-	-	-	-	-	-
Temperament	-	1.80	2.45	5.14	0.099	-	-	-	-	-	-	-	-	-	-
Plumage colour	3.06	1.51	7.19	11.76	0.227	1.83	1.74	3.11	6.68	0.021	-	1.37	4.48	5.85	0.018
Chicken shape	1.17	3.49	12.17	16.83	0.325	-	-	-	-	-	-	-	-	-	-
Meat quality	4.34	9.94	2.89	16.57	0.319	11.50	9.76	13.35	34.61	0.094	8.63	6.69	18.49	33.80	0.099
Egg size	12.58	5.06	3.10	20.74	0.400	34.33	13.59	1.24	49.16	0.192	37.45	12.46	-	49.91	0.205
Egg shell colour	-	-	-	1.26	0.024	11.83	30.49	3.73	46.05	0.172	12.26	31.91	4.20	48.38	0.180
Egg yolk colour	-	-	-	1.38	0.027	-	-	-	1.17	0.010	1.12	-	1.36	2.48	0.010

Source: Okeno *et al.*, 2012.

The aim of this study is to identify signatures of selection in production genes for eggs and growth in poultry. The molecular approach is an accurate method of selecting preferred traits compared to traditional selection methods (Dekkers, 2002). This is because the genetic architecture of a phenotype is known compared to traditional selection method that relies on observation of the phenotype which may be influenced by external factors such as the environment. (Naqvi, 2007). This molecular approach reduces the generation interval as DNA is available at any stage of the animal unlike in traditional selection where the phenotype is recorded at a certain stage. Using the molecular approach in addition to traditional selection methods may lead to success in improving poultry production (Dekkers, 2002).

Agriculture accounts for 25-26% of the Gross Domestic Product (GDP) in Kenya with poultry farming contributing 30% of agricultural contribution to the country's GDP (Nyaga, 2007).

Indigenous poultry can thus alleviate poverty and the Millenium Development Goals (MDG) of halving the number of poor people and empowering women will be met, if high producing genotypes are also identified (Moreki *et al*., 2010; FAO, 2014).

1.3 Statement of the problem

Production traits have not been successfully improved using the traditional selection methods. These methods are based on the use of phenotypic selection of traits that are observable and measurable (Naqvi, 2007). Animals with superior performance are used as the breeding stock (Naqvi, 2007). The traditional methods are difficult to improve traits for egg and meat production which have low heritability.

Improving these traits with low heritability is also costly as the phenotype is measurable only later in life (Dekkers, 2002). In this case, a large number of animals have to be raised so that a few are selected for breeding.

It is also difficult to improve traits that are not easily measurable like those which are recorded in one sex only or those that can be recorded after the death of an animal using the traditional selection methods (Dekkers, 2002; Naqvi, 2007). Crossbreeding of indigenous poultry with exotic genotypes is also used which leads to genetic erosion of the indigenous genotypes (FAO, 2014). The purpose of this study was to use the *in silico* approach to identify signatures of selection in production genes for egg and meat. This may be used as an additional tool to phenotypic selection by the government and other stakeholders to enhance selection by identifying superior breeds for the genetic improvement of economic traits in poultry (Dekkers, 2002).

1.4 Justification

As the computation approach uses genetic information, there is higher heritability because genetic information is not influenced by the environment (Naqvi, 2007). The use of genetic information allows for early selection and the generation intervals are reduced since the molecular information is available as early as at the embryo stage (Dekkers, 2002). Genetic information is available on all selection candidates: sex-limited traits, traits which are difficult to measure, or traits which require the animal to be slaughtered (Dekkers, 2002). As the use of this approach enhances selection intensity and accuracy with reduction in generation intervals, the rate of genetic change is high (Dekkers , 2002).

Using the molecular approach in addition to traditional selection methods may lead to success in improving poultry production
(http://www.ccsi.ca/Meetings/canmore/jack_dekkers). This will benefit the resource-limited farmers who form the majority of poultry farmers.

Poultry farming has the potential to alleviate poverty (Moreki *et al*., 2010). When poultry production is improved, the economy of the country will also improve as poultry contributes to 30% of agricultural contribution to the country's GDP (Nyaga, 2007).

1.5 Research hypothesis

Null hypothesis: There are no signatures of selection in genes for egg production and growth in poultry.

Alternative hypothesis: There are signatures of selection in candidate genes for egg production and growth in poultry.

1.6 Objectives

1.6.1 General objective

To perform computational analysis of molecular signatures of selection in candidate genes involved in egg production and growth in poultry.

1.6.2 Specific objectives

1. To analyze and identify molecular signatures of selection in Prolactin, Vasoactive intestinal peptide 1, and vasoactive intestinal peptide receptor 1 genes implicated in egg production in poultry.

2. To analyze and identify molecular signatures of selection in Growth hormone, Growth hormone receptor, Insulin-like growth factor 1, and Insulin-like growth factor 1 receptor genes implicated in growth in poultry.

3. To predict secondary and tertiary structures of proteins encoded by the selected candidate genes for growth and egg production in poultry.

4. To compare the putative 3D structures of the proteins encoded by the candidate genes between the homologs.

CHAPTER TWO

LITERATURE REVIEW

2.1 Candidate genes and their functions

2.1.1 Prolactin

Avian species have a single prolactin-encoding gene which is located on chromosome 2 (Alipanah *et al*., 2011). The prolactin gene has been cloned in different avian species, examples including turkey, quail, duck, pigeon, and domestic chicken. The avian prolactin gene consists of five exons and four introns (Yousefi *et al*., 2012). The gene has two regulatory regions, the proximal and distal enhancers. The distal segment is responsible for most of the transcription activity. Expression of the avian prolactin gene occurs in the ovary, oviduct, and hypothalamus. The highest expression is found in the pituitary gland (Li *et al*., 2009).

The prolactin hormone is the expressed product of the prolactin gene. It is composed of 199 amino acids (Wilkanowska *et al*., 2013). The hormone is a polypeptide which is secreted by the anterior pituitary gland and by other numerous tissues and cells (Freeman *et al*., 2000). It has more than 300 functions in various vertebrates (Bole-Feysot *et al*., 1998). The functions of prolactin can be divided into the following categories: 1) Water and electrolyte balance 2) Growth and development 3) Endocrinology and metabolism 4) Brain and behavior 5) Reproduction

6) Immunoregulation and protection.

In avian species, the induction of broodiness and incubation is one of the most important functions of prolactin (Jiang *et al*., 2005). Riddle *et al*., (1935) demonstrated that injection of prolactin induces incubation behavior in birds. However, changes in the promoter binding site of prolactin due to polymorphism influences mRNA expression conversely influencing incubation behavior and egg production. For example, an insertion of a 24 bp at the promoter region of prolactin is associated with egg production (Cui *et al*., 2006 ; Usman *et al*., 2014). To perform any of its functions, prolactin interacts with its receptor in various target cells leading to a cascading activation of intracellular events (Freeman *et al*., 2000).

The prolactin-receptor binding happens in a two-step process: 1) one site of prolactin binds to one receptor molecule 2) a second receptor molecule binds on a second site of prolactin (Bole-Feysot *et al*., 1998, Freeman *et al*., 2000). This forms a homodimer of prolactin and the two receptor molecules. The prolactin associates with JKA2 which is a cytoplasmic tyrosine kinase (Freeman *et al*., 2000). On dimerization, tyrosine phosphorylation is induced followed by activation of the JAK kinase and phosphorylation of the receptor. The phosphorylation of cytoplasmic stat proteins, which dimerize and translocate to the nucleus then bind specific promoter elements on PRL-responsive genes is a major pathway of signaling. The PRL also activates the Ras/Raf/MAP kinase pathway which may be involved in the proliferation effects of the hormone. IRS-1, P1-3 kinase, SHP-2, PLCγ, PKC and intracellular Ca^{2+} are also potential mediators (Bole-Feysot *et al*., 1998).

2.1.2 Vasoactive intestinal peptide 1 and vasoactive intestinal peptide receptor 1

The vasoactive intestinal peptide 1 (VIP1) gene is 1,069 bp long. The avian VIP1 gene is found on chromosome 3 (https://www.ncbi.nlm.nih.gov/nuccore/NM_205366.2).

The expressed protein is vasoactive intestinal peptide 1 (VIP1) which is a 28 amino acid, regulatory peptide whose actions are through its interaction with specific receptors which are coupled to adenylyl cyclase and the production of cAMP (Igarashi *et al*., 2011). It has regulatory effects on the circulatory, reproductive, immune and gastrointestinal systems. Chicken VIP1 is widely distributed in the peripheral organs and the brain. It is found notably in the endocrine glands where it exerts pleiotropic physiological functions.

 In avian species, it stimulates prolactin release although it is not the main regulator and also stimulates prolactin gene expression (Chaiseha *et al*., 2004). In the avian reproductive cycle, VIP1 controls prolactin (PRL) secretion at the hypothalamus while vasoactive intestinal peptide receptor 1 (VIPR1) partly controls it in the pituitary (Chaiseha *et al*., 2004).

It also occurs in high concentrations in the hypothalamus. A study done by Tong *et al*., (1998) on turkey hens showed that VIP1 regulates the expression of PRL gene by increasing its concentration by acting both at the transcriptional level and at the level of PRL mRNA stability. On the other hand, El-Halawani *et al*., (2000) demonstrated the action of VIP1 on incubation behavior through the active immunization of turkey hens against VIP1 which resulted in inhibition of incubation behavior and subsequent

increase in egg production. The role of VIP1 is regulated through the VIPR1 gene expression in the pituitary (Zhou *et al*., 2008).

The chicken vasoactive intestinal peptide receptor 1 (VIPR1) gene is 67,906 bp long (Zhou *et al*., 2008). It is located on p3.2 of chromosome 2 (Kansaku *et al*., 2001).

The chicken VIPR1 gene is composed of 13 exons. The exons range in size from 45 to 1,031 bp. The VIPR1 gene is expressed in the pituitary and the hypothalamus. The differential mRNA expression of the VIPR1 gene is associated with changes in reproduction (Chaiseha *et al*., 2004).

The VIPR1 is a glycoprotein that belongs to the class II subfamily of the 7-transmembrane G-protein-coupled receptors superfamily (Zhou *et al*., 2008). It has a large hydrophilic extracellular N-terminus followed by 7 highly conserved hydrophobic transmembrane helices and a cytoplasmic C- terminus. In avian species, VIPR1 is found in the hypothalamus, on the surface membranes of the anterior pituitary, the small intestines and the granulosa cells (Zhou *et al*., 2008). DNA variation of VIPR1 is associated with broodiness in avian species (Zhou *et al*., 2008).

2.1.3 Growth hormone and growth hormone receptor

The chicken growth hormone (GH) gene is about 4 kbp (Tanaka *et al*., 1992). It is located on the long arm of chromosome 1q4. The chicken GH gene consists of five exons and four introns (Tanaka *et al*., 1992). Although the chicken GH gene is analogous to mammalian GH genes, the size is significantly larger. The intron size of the

chicken GH gene makes it bigger by 3.5 kbp. Primer extension analysis determined the start point of transcription of the GH gene to be 56 bp upstream from the start codon (Tanaka *et al*., 1992).

Growth hormone (GH) is a 190 amino acid protein that is synthesized and secreted in the anterior pituitary gland (Buggiotti and Primmer, 2006). It performs many physiological actions among which is controlling growth by regulating insulin-like growth factor 1 (IGF1) concentrations, stimulating lipolysis, providing FFAs, and glycerol as substrates for energy metabolism and also inhibiting insulin-induced suppression of hepatic gluconeogenesis (Clemmons, 2004).

The growth hormone also maintains carbohydrate and lipid homeostasis under normal conditions. Many of these functions are exerted by regulation of the transcription of other genes including IGF1 transcription factors and metabolic enzymes (Clemmons, 2004).

On the other hand, the chicken growth hormone receptor (GHR) gene is located on chromosome Z (https://www.ncbi.nlm.nih.gov/gene/408184). Chicken GHR gene has 9 exons.

Growth hormone receptor (GHR) belongs to the vertebrate somatotropic axis which regulates post natal growth and metabolism (Varvio *et al*., 2008). It is also a member of the large cytokine receptor superfamily and it has 3 domains which are an extracellular ligand-binding domain, a transmembrane domain and a cytoplasmic domain. Growth hormone binds to the extracellular domains of the two GHR molecules leading to GHR

dimerization, activation of the GHR-associated JAK2 (Janus-Family Tyrosine Kinase-2), and tyrosyl phosphorylation of both JAK2 and GHR (Frank, 2001). This leads to activation of a variety of signaling molecules including MAPKs (Mitogen-Activated Protein Kinases), IRS1 (Insulin Receptor Substrate-1), PI3K (Phosphatidylinositol- 3-Phosphate-Kinase), DAG (Diacylglycerol), PKC (Protein Kinase-C), Ca^{2+} (intracellular calcium), and STATs (Signal Transducers and Activators of Transcription).

The signaling molecules contribute to the changes in enzyme activity, transport function and gene expression all of which are GH-induced. This ultimately leads to changes in growth and metabolism (Varvio *et al*., 2008).

2.1.4 Insulin-like growth factor 1 and insulin-like growth factor 1 receptor

The insulin-like growth factor 1 (IGF1) gene is 50 kbp long (Kajimoto and Ratowein, 1991). Genetic linkage analysis has determined the chicken IGF1 gene to be located on the short arm chromosome 1 near the centromere (Klein *et al*., 1996). The expression of IGF1 gene is ubiquitous including in the kidney, testes, liver, muscle, heart, ovary, brain, intestines, among other tissues (Tanaka *et al*, 1996).

IGF1 plays a role in various physiological and metabolic processes for example stimulating myogenesis, inhibiting apoptosis, increasing the absorption of glucose, activation of cell cycle genes and it also intervenes in the synthesis of DNA, RNA, protein and cell proliferation (Etherton, 2004).

The IGF1 gene is hence used as a candidate gene in the prediction of growth and meat quality traits in genetic improvement programs (Andrade *et al*., 2008). The fundamental role of IGF1 in prenatal and postnatal development and other physiological roles are through the binding to IGF1 receptor (Delafontaine *et al*., 2004).

On the other hand, the insulin-like growth factor 1 receptor (IGF1R) gene is 10,818 bp long. It is located on chromosome 10. Chicken IGF1R has 21 exons (https://www.ncbi.nlm.nih.gov/gene/395889).

The insulin-like growth factor 1 receptor (IGF1R) protein belongs to the receptor tyrosine kinase family (Delafontaine *et al*., 2004). It is phosphorylated following the binding to IGF1 (Delafontaine *et al*., 2004). IGF1R is a large transmembrane glycoprotein dimer which consists of several structural domains. The N terminal of the ectodomain contains two leucine-rich repeat domains separated by a cysteine-rich region while the C-terminal of the ectodomain consists of three fibronectin type III domains (Adams *et al*., 2000). The IGF1R is widely expressed in many cells in the fetal and postnatal tissue where it stimulates growth. IGF1R binds IGF1 and also IGF2 although with a lower affinity. IGF1, IGF2 and insulin ligands share a common 3D architecture and can bind IGF1R in a competitive manner. On activation, the IGF1R functions in cell growth and survival control.

When the IGF1R binds with its ligand, the receptor kinase is activated leading to receptor autophosphorylation, and tyrosine phosphorylation of multiple substrates which function as signaling adapter proteins including, the insulin-receptor substrates (Insulin

receptor substrate (IRS)1/2), Shc and 14-3-3 proteins (Delafontaine *et al*., 2004). The phosphorylation of IRSs proteins leads to the activation of two main signaling pathways: the PI3K-AKT/PKB pathway and the Ras-MAPK pathway. Activation of the MAPK pathway leads to increased cellular proliferation, whereas activating the PI3K pathway inhibits apoptosis and stimulates protein synthesis. Phosphorylated IRS1 can activate the 85 kDa regulatory subunit of PI3K (PIK3R1), leading to activation of several downstream substrates, including protein AKT/PKB. AKT phosphorylation, in turn, enhances protein synthesis through mTOR activation and triggers the anti-apoptotic effects of IGFIR through phosphorylation and inactivation of BAD.

IPI3K-driven signaling leads to recruitment of Grb2/SOS by phosphorylated IRS1 or Shc leads to recruitment of Ras and activation of the ras-MAPK pathway. IGF1R also signals through the Janus kinase/signal transducer and activator of transcription pathway (JAK/STAT). Phosphorylation of JAK proteins can lead to phosphorylation/activation of signal transducers and activators of transcription (STAT) proteins. Particularly, STAT3 activation may be essential for the transforming activity of IGF1R. The JAK/STAT pathway activates gene transcription and may be responsible for the transforming activity. JNK kinases can also be activated by the IGF1R. The activation can be inhibited by IGF1 through phosphorylation and inhibition of MAP3K5/ASK1, which is able to associate with the IGF1R directly. When present in a hybrid receptor with INSR, it binds IGF1 (Delafontaine *et al*., 2004).

2.2 Gene signatures of selection

Variation of genes in organisms both within and between species comes about by several evolutionary forces among which are natural selection, single nucleotide polymorphism, recombination, genetic drift and migration among others. This genetic polymorphism may lead to phenotypic polymorphism.

Natural selection

Natural Selection is the process by which some phenotypes adapt to their environment, survive and reproduce (Aquadro *et al*., 2001).

There are three forms of natural selection (Nielsen, 2005) i) positive selection which favors alleles that are advantageous   ii) neutral/balancing selection which favors both alleles in a population hence maintaining polymorphism iii) negative/purifying selection which eliminates mutations which are unfavorable.

The different modes of selection are a response to the external pressure with each leading to change in allele frequency (Oleksyk *et al*., 2010). Positive selection on favoring advantageous alleles leads to a decrease in genetic variation. On the other hand, purifying selection on eliminating deleterious mutations maintains the integrity of functional sequences while balancing selection acts to maintain polymorphism (Oleksyk *et al*., 2010).

There are several postulations on the nature of variation among living organisms (Nielsen, 2005). According to the neutral theory, a large proportion of variation both within and between species is selectively neutral.

Meaning that new mutations that arise do not confer a fitness advantage as the loci are not affected by selection (Kimura, 1983). New mutations (due to random factors) which may not confer any fitness advantage may occur and increase in frequency in the population. The process by which the frequency of alleles change in a population due to random factors is known as genetic drift (Nielsen, 2005). Another postulation maintains that most of the variation observed does affect the fitness of the organism and is subject to positive selection (Gillespie, 1991). There is however an unsettled debate on the true postulation (Nielsen, 2005).

The regions with DNA sequences which are involved in variation of phenotype due to natural or artificial selection are called selection signatures (Qanbari *et al*., 2012). Each mode of selection leaves a specific signature on the genome architecture (Oleksyk *et al*., 2010).

The search for molecular signatures of selection is motivated by: 1. Interest in species' evolutionary past and the mechanism of the evolution 2. The association of the signatures with traits of interest has adaptive or functional importance (Nielsen *et al*., 2007).

The studies are possible due to development of various tools and methods (Lopez *et al*., 2015). Negative and purifying selection may be of great interest because it may help

detect regions or residues of functional importance, but positive selection generates more interest because it is associated with formation of new functions and adaptation (Nielsen, 2005). When alleles that have undergone positive selection increase in a population, they become fixed in the population bringing about genetic variation and subsequent phenotypic variation.

The signatures of positive selection are divided into five main groups:

2.2.1 Long haplotype

A positively selected allele can become prevalent rapidly such that recombination does not substantially break the association with alleles at nearby loci on the ancestral chromosome (Sabeti *et al*., 2006). This collection of alleles in a chromosomal region that occur together is called a haplotype. Selective sweeps may produce an allele with high prevalence and long range associations with other alleles. The long range associations are a long haplotype which have not been broken down by recombination (Sabeti *et al*., 2006). This signature detects positive selection that occurred <30,000 years ago (Sabeti *et al*., 2006). The method used to detect this is the LD and (Long-range haplotype) LRH (Toomajian *et al*., 2003).

2.2.2 Population differentiation

This signature detects positive selection that happened between <50,000-75,000 years ago (Sabeti *et al*., 2006).

Population differentiation can only arise when populations living in separate geographical regions are subject to different environmental or cultural pressure, positive selection may change the frequency of an allele in one of the populations (Sabeti *et al*., 2006). Positive selection may lead to genetic variation in one population and not the other. Indices used to detect this include the breeding coefficient F_{ST} (Akey *et al*., 2002).

2.2.3 Reduction in genetic diversity

When an allele increases in population frequency, variants at nearby locations on the same chromosome also rise in frequency. This so called "hitchhiking" leads to a selective sweep that alters the genetic variation pattern of the region (Sabeti *et al*., 2006). When a complete sweep occurs, the allele that has been selected together with the closely linked variants becomes fixed eliminating diversity. This diversity is restored when new slow mutations occur. These selective sweeps are detected in genetic variation within species. SNP is the most commonly used variant. This signature is used to detect positive selection that happened <250,000 years ago (Sabeti *et al*., 2006). Statistical tests used for this detection include deviations from HWE, Tajima's D, HLA test and Fu and Li's D and D*, Fu and Li's F and F*, heterozygosity and Ewens Watterson homozygosity (F) (Hudson *et al*., 1987).

2.2.4 High frequency derived alleles

Derived or non-ancestral alleles arise as a result of new mutation and usually have lower allele frequencies compared to ancestral alleles (Watterson and Guess, 1977). Derived alleles that are linked to a beneficial allele can hitchhike to high frequency in a selective

sweep (Sabeti *et al*., 2006). Many of the derived alleles will not reach complete fixation thus a signature of a region that contains many high-frequency derived alleles is created by positive selection.

This signature detects positive selection that occurred <80,000 years ago (Sabeti *et al*., 2006). The test commonly used to detect this is the Fay and Wu's H (Fay and Wu, 2000).

2.2.5 High proportion of function altering mutation

Molecular variations that alter the function of a protein are deleterious and are less likely to become fixed compared to mutations that do not lead to a change in protein function (Sabeti *et al*., 2006). Fixation of function altering mutations that are beneficial can be increased when positive selection occurs over a prolonged period (Li *et al*., 1985).

The changes can be measured by comparing DNA sequences between species. Detection of the increase may be done by comparing the rate of non-synonymous to the rate of synonymous mutations (Li *et al*., 1985). This signature can be determined using tests such as ka/ks ratio, McDonald and Kreitman tests (MK tests), d_N/d_S ratios and codon substitution models implemented in the PAML package (McDonald and Kreitman 1991). This signature is studied the most and is used to detect adaptive evolution.

The d_N/d_S ratio is the most popular test to model positive natural selection in the gene sequences in terms of positive adaptive selection (Ngandu *et al*., 2008). This is the relative rate of non- synonymous to synonymous substitutions in the evolutionary history of a sample of sequences.

The d_N/d_S ratio can detect positive selection between and within species. The codon substitution models of Nielsen and Yang (1998), as implemented in the PAML package promoted d_N/d_S to the status of a parameter which they called omega (ω). Values of omega <1 indicate negative or purifying selection while values of omega >1 indicate positive or diversifying selection (Yang, 2007). When the value of omega is equal to 1, this is an indication of neutral selection (Yang, 2007).

2.3 Lineage specific positive selection

An approach for detecting positive selection is by use of comparative data that detects positive selection within and between species. For lineage specific positive selection, two codon based models are commonly used, Model 0 which is the one-ratio model and the free-ratio model (Model 1). The one-ratio model assumes an equal $\acute{\omega}$ ratio for all the branches in the phylogeny while the free-ratio model assumes an independent $\acute{\omega}$ ratio for each branch (Yang, 1998).

The two models are then compared by the likelihood ratio test (LRT) using the following formula: $LRT = 2 (lnLM_0 - lnM_1)$ (Yang, 1998). The result is then compared to a χ^2 distribution with N-1 degrees of freedom where N is the number of branches. A significant difference in the LRT value is an indication that the $\acute{\omega}$ results obtained are reliable (Yang, 1998).

2.4 Site specific positive selection

For site specific positive selection, again two codon based models are commonly used, the model 7 (beta) which uses a beta distribution hence allows sites with ω values of 0 to 1 and Model 8 (beta and ώ) adds an extra class of sites to model 7 with the proportion and the ώ ratio estimated from the data hence allowing for sites with ω >1 (Yang *et al.*, 2000). The two models are also compared by the LRT i.e. LRT = 2 ($\ln LM_7$ - $\ln M_8$). For the sites specific models, posterior probabilities are also calculated for sites under positive selection and values >0.95 indicate that the results for the amino acid sites under selection are reliable (Nielsen & Yang, 1998; Yang *et al.*, 2005).

Few studies have been done to detect positive selection at candidate genes that influence production of eggs and meat in livestock. When positive selection is detected on the genes that influence the general production of eggs and meat, the phenotype can be confirmed by carrying out *in vitro* and *in vivo* studies.

Zhu *et al.*, (2010) used the codon substitution models to identify positive selection in the mx gene of the galliform and anseriform avian. The mx proteins confer different antiviral activity in many different species (Zhu *et al.*, 2010). In the study, the positively selected sites were distributed in the N-terminal end of the mx gene and GTPase domain which are important for the antiviral activity of mx protein. These may be used as molecular markers for breeding poultry with improved antiviral activity (Zhu *et al.*, 2010).

2.5 Protein 3D structure prediction

The structure of a protein is valuable in determining the functions of a protein like enzymatic activity or the interaction with other proteins (Bairoch, 2000). The solving of protein structures experimentally such as those in Protein Data Bank (PDB) is time consuming and costly and can be sometimes technically difficult (Kallberg *et al*., 2012). There are different computational protocols that have been developed to solve the problem of the many protein sequences that have been deposited without solved structures. There are three categories of computational structure prediction methods: Comparative/template modeling, threading and template free/*ab initio* modeling (Kallberg *et al*., 2012).

In comparative modeling, the structure is constructed by aligning the sequence or sequence profile of the target protein to an evolutionary related protein which has a solved structure (template) to find equivalent residues (Marti-Renom *et al*., 2000). In threading, the target sequence is matched to a solved 3D structure template which has no evolutionary relationship (Jones *et al*., 1992). Threading is based on the principle that evolutionary distant or protein sequences that are unrelated often share common structural elements. Structure predictions based on existing template structures like comparative modeling and threading produce high resolution models. However, these models are only useful if close templates are available (Kallberg *et al*., 2012). When there are no solved structures for homologs, *ab initio* methods construct models without using such information. However, this method can successfully create models from sequences which are <120 residues (Wu *et al*., 2007).

Accuracy of the predicted model is important for it to be biologically useful. The accuracy is largely dependent on the availability of close structural templates and the prediction method.

The RaptorX is a software that can create models in which there are no close templates (Peng & Xu, 2011). It optimizes the modeling strategy by using the profile entropy scoring method that considers the number of non-redundant homologs which are available for the target and the template structure to assess the information in sequence profiles (Peng & Xu, 2010). The software also integrates biological signals in a nonlinear score function using conditional random fields (CRFs) (Peng & Xu, 2009).

In addition, RaptorX uses a Multiple-Template Threading (MTT) procedure which uses multiple templates to model a target sequence (Multiple-Template Threading) procedure (Peng & Xu, 2011). It is possible to have many good templates for a given target which may improve the modeled structure. Using the MTT, the alignment between individual pairwise sequence-template alignments is improved through the use of a probabilistic consistency transformation (Källberg et al., 2012).

Thiel et al. (2015) used Raptor X to predict the structure of CD45 in primates. CD45 has a crucial role in the immune system where it is central to the development of T cells and also determines T and B lymphocytes activation (Thiel et al., 2015). Viruses interfere with the immune system through the interaction of viral proteins with the CD45. This interaction is the driving force for the evolution of CD45 (Thiel et al., 2015). In this

study, they modeled the ectodomain of the CD45 which consists of three fibronectin III-like repeats and a cysteine-rich domain.

Sites under positive selection were distributed in the different domains. The sites were found at the surface of the ectodomain where pathogens could easily interact with the CD45 through the viral proteins. This adaptive selection is important to the host in that it prevents the virus from evading the immune system (Thiel *et al*., 2015).

2.6 Protein structure alignment

A protein structure alignment method consists of two major components: a scoring function measuring protein similarity and a search algorithm optimizing the scoring function (Wang *et al*., 2013).

There are different programs for automatic structure alignment whose accuracy is low compared to manually-curated alignments which consider the geometric similarity, evolutionary and functional relationship (Mayr *et al*., 2007). Most of these programs produce alignments that are not biologically meaningful. DeepAlign is one of the programs that produces biologically meaningful alignments as its scoring function considers the 3D geometric similarity, evolutionary information at the sequence and local substructure levels as well as hydrogen-bonding similarity (Wang *et al*., 2013). Proteins in a family share an overall shape but their structures are highly flexible due to evolution at the local substructure and the sequence level. DeepAlign uses amino acid and local substructure substitution matrices to quantify the local conformational change resulting from evolutionary events (Wang *et al*., 2013). Block substitution matrix

(BLOSUM) is the commonly used substitution matrix. It is not sensitive for remote homologs as it is derived from close homologs.

DeepAlign hence uses conformation letter substitution matrix (CLESUM) which is a local substitution matrix to measure the evolutionary distance of two proteins at the local substructure level (Zheng, 2008; Wang *et al*., 2013). In addition, to make alignments that are more biologically meaningful especially for beta-strands, DeepAlign uses the hydrogen-bonding similarity. Conformation letter substitution matrix (CLESUM) favors the alignment of two loop regions that are evolutionary related but disfavors that of unrelated helices.

To reduce the running time used to find an optimal structure alignment, DeepAlign identifies a set of initial alignments. It uses both amino acid and local substructure mutation scores to identify evolutionary related structural fragment pairs then builds initial alignments from these pairs (Wang *et al*., 2013).

In order to generate biologically meaningful structural alignments, all four factors are necessary (Wang *et al*., 2013). However the 3D geometric similarity which includes the TMscore, RMSD and uGDT is the major factor for determining similarity of proteins. TMscore is better than RMSD as it is independent of the length of the protein and is not biased by only a few residues that are badly aligned (Wang *et al*., 2013). Structure alignments with a TM value of > 0.5 are mostly in the same fold while those with TM<0.5 are not in the same fold (Xu & Zhang, 2010).

CHAPTER THREE

MATERIALS AND METHODS

3.1 Retrieval of candidate gene sequences

The full coding sequence mRNA sequences of chicken prolactin, vasoactive intestinal peptide 1, and vasoactive intestinal peptide receptor 1 implicated for egg production and growth hormone, growth hormone receptor, insulin-like growth factor 1, insulin-like growth factor 1 receptor implicated for growth were retrieved from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) using ENTREZ searches across the databases. The reference genome used was the chicken genome available at ENSEMBL (Hubbard *et al*., 2002).

3.2 Search for homologs and retrieval

A reciprocal BLAST was done using the algorithm BLASTp which is available at NCBI to infer homology. The amino acid sequence of each candidate gene in the fasta format was used as the query in performing pairwise sequence alignments in non-redundant (nr) databases. The substitution matrix BLOSUM62 was used. For egg production, homologues selected were birds and amphibians with an E value greater than 1e-10. For meat production, the homologues selected were animals that are commonly eaten by man with an E value of greater than 1e-10. Amino acid sequences of the homologues were converted to their corresponding coding sequences and the stop codons manually removed to prevent interference with later analysis.

3.3 Multiple sequence alignment

ClustalX2 (Larkin *et al*., 2007) which is a Multiple sequence alignment software that uses a progressive algorithm which is heuristic in nature was used to align the different homologues for each candidate gene to confirm homology and as a preceding step in phylogenetic analysis. PRANK that also uses the progressive algorithm and MAFFT (Katoh & Standley, 2013) were used. This was followed by multiple sequence alignment using MUSCLE version1.3.8.31-1 (Edgar, 2004). Jalview version 2.8 (Waterhouse *et al*., 2009) and Seaview were used to view and edit the alignments. On comparing of the alignments from the three softwares, alignments from MUSCLE were found to be better and were used for subsequent analysis.

3.4 Phylogeny construction

MEGA6 was used to model the amino acid substitution and the rate heterogeneity within the various homologues (Tamura *et al*., 2013). The model was selected using the Lowest Bayesian Information Criterion. Phylogenetic trees were then inferred using Nearest Neighbour Interchange with subtree pruning and regrafting of FastME2 (Lefort *et al*., 2015). 1000 bootstraps were used to test for confidence of the inferred relationships. The inferred trees were saved in Newick format. Figtree version 1.4.2 (Rambaut, 2014) was used for graphical visualization of the trees.

3.5 Substitution model selection and rate heterogeneity testing

Codon-based substitution models which were developed by Nielsen and Yang (1998) and Yang (2007), which are implemented in CODEML program of PAML were used to detect selection in lineage and in sites.

For lineage, 2 codon-based models: Model 0 or one ratio Model and Model 1 also known as the free ratio Model were used. The one ratio model assumes an equal ω ratio for all branches in the phylogeny. The free ratio model assumes an independent ω ratio for each branch. For the site specific, 2 codon-based models were used: Model 7(beta) and Model 8(beta and ω). Model 7 allows for sites with values ω of 0 to 1 while Model 8 allows for sites with ω >1.

3.6 Input preparation in PAML

The Perl script copygaps_single.pl was then used to align the aligned protein sequence with the corresponding codon sequences. The output file for the analysis was saved as infile.nuc. At the UNIX prompt, three directories were created: Model0, Model1 and Sites. The codeml.ctl file was edited according to each directory corresponding to the analyses that were to be performed.

For Model0 directory, Model 0 was selected, for Model 1 directory, Model 1 was selected while for the Sites directory, Model 7 and 8 were selected. The phylogeny tree that was saved in Newick format was then moved into the intree.trees file. The

codeml.ctl which is the control file, infile.nuc which is the output from the Perl script and intree.trees which contained the tree file were each copied into the three directories.

3.7 Test for molecular signatures of selection

The PAML5 package (Yang, 2007) was then used for phylogenetic analysis using maximum likelihood. CODEML was run on the command prompt. This was done in two parts, first was the branch or lineage analysis and the second was the amino acid or site analysis. After the analysis, the log likelihood ratio test (LRT) (Yang, 1998) was used to compare two nested models: a null model that does not allow for any codons with ω ratio of greater than 1 against a more general model that does. Twice the log likelihood difference between the two models was compared against a chi squared (X^2) distribution to test the significance of the result obtained. For Lineage analysis, LRT was also used to compare Model7 (beta) with ω ratios of between 0 and 1 and Model8 (beta & ω) that adds a class of sites with dN/dS of >1 were used. Bayes Empirical Bayes analysis was done to infer the category an amino acid site belongs to (Yang *et al*., 2005).

Amino acid sites with a predetermined posterior probability of >0.95 belong to a site class that has a dN/dS ratio of > 1 and thus believed to be under positive selection (Yang *et al*., 2005).

3.8 Protein 3D structure prediction

The secondary and tertiary structures of the proteins encoded by the candidate genes for growth and egg production for poultry were modeled by sending the sequences to

RaptorX (http://raptorx.uchicago.edu/) which is a webserver that models protein structures using the homology algorithm. Pymol version 1.7.6 (DeLano, 2006) was used to view the 3D structures of the proteins.

3.9 Comparative 3D structure analysis

The sequences for the genes for egg production and growth for poultry were submitted to DeepAlign (http://raptorx.uchicago.edu/DeepAlign) which is a server for comparative structure analysis.

CHAPTER FOUR

RESULTS

4.1 Homology search

From the reciprocal BLASTp performed in the non-redundant databases, the orthologs (shown in Appendix 1-8) were selected. These had an E value lesser than 1e-10 for the different genes.

4.2 Analysis of signatures of selection at candidate genes for egg production in poultry

4.2.1 Multiple sequence alignment of prolactin

On using Muscle software to perform the Multiple sequence alignment of prolactin homologous, the alignment in Figure 4.1 was produced. The domain for prolactin is highly conserved across the poultry (greylag goose, wild duck, wild chicken, grey jungle fowl, Japanese quail, common quail, chicken breed lohmann, chicken breed yangshan, chicken breed 15 Hubbard, chicken cob 500, chicken breed avian 48) with a few substitutions. However, there were notable amino acids substitutions in common quail, Japanese quail, helmeted guinea fowl. In position 24 of common quail and Japanese quail, serine was substituted by glycine. In position 26 of common quail, arginine was substituted by threonine. In addition, in position 124 of common quail, Japanese quail, and common quail, glycine was substituted by alanine, in position 171 of common pea fowl, serine was substituted by alanine. For helmeted guinea fowl, threonine substituted

arginine in position 213. In position 253 of Japanese quail and common quail, glycine substituted alanine.

Figure 4.1: Multiple sequence alignment of prolactin protein. The alignment shows the degree of identity between the species. The sequences of different poultry species are highly conserved with few substitutions.

4.2.2 Multiple sequence alignment of vasoactive intestinal peptide 1 (VIP1)

On doing a Multiple sequence alignment of the VIP1 homologs using the Muscle software, the alignment below was produced (Figure 4.2). The domain for VIP1 is also conserved in different poultry species (ostrich, wild turkey, southern ostrich, swan

goose, wild chicken) as shown in Figure 4.2. However, this sequence has many indels that are widely spread across the sequence even in the poultry species. Interestingly, the ostrich sequence is highly variable compared to the other poultry sequences.

In position 159 of the ostrich sequence, aspartate substituted asparagine, alanine substituted threonine in position 186, valine substituted proline in position 188, aspartate replaced asparagine in position 190, glycine substituted serine in position 253, glycine replaced arginine in position 241 and glutamate was replaced by glycine in position 293.
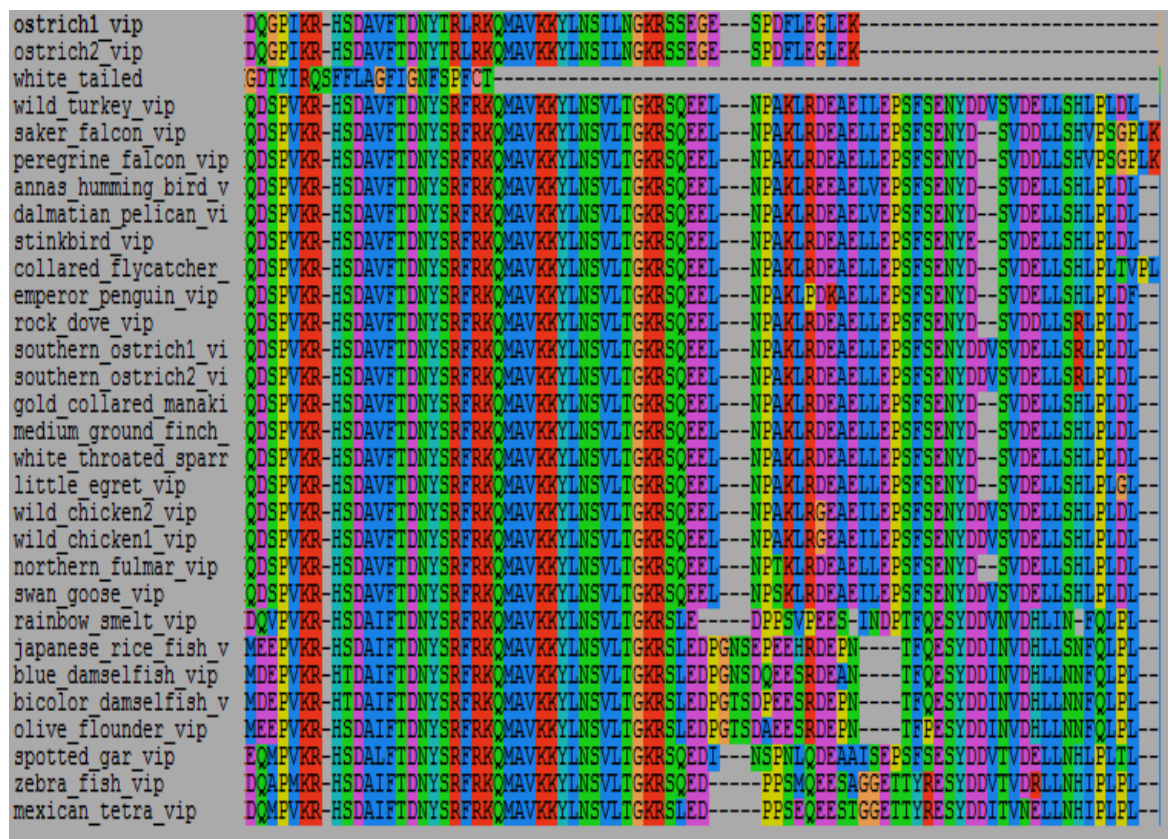
Figure 4.2: Multiple sequence alignment of vasoactive intestinal peptide 1 (VIP1) protein. The alignment shows that there have been many insertions/deletions that have occurred. The sequence of the ostrich is highly variable from other poultry sequences.

36

4.2.3 Multiple sequence alignment of vasoactive intestinal peptide receptor 1 (VIPR1)

From the Multiple Sequence Alignment of VIPR1 homologous sequences done using Muscle software, the alignment (shown in Figure 4.3) was produced.

VIPR1 gene is conserved with minimal variation within the poultry (wild turkey, wild duck, wild chicken, Japanese quail and Southern ostrich). The Southern ostrich and the wild duck are the most varied of the poultry species.

In the Southern ostrich, there are 6 sites with substitutions: at position 131, glutamic acid substituted glutamine, at position134, alanine was substituted by glutamic acid, at position 151, tyrosine substituted phenylalanine, at position 256, serine substituted proline and at position 512, glycine substituted serine. The wild duck had several substitutions at position 183 where phenylalanine substituted tyrosine, at position 196, serine substituted alanine, at position 361, threonine was substituted by alanine and at position 428 threonine replaced alanine.
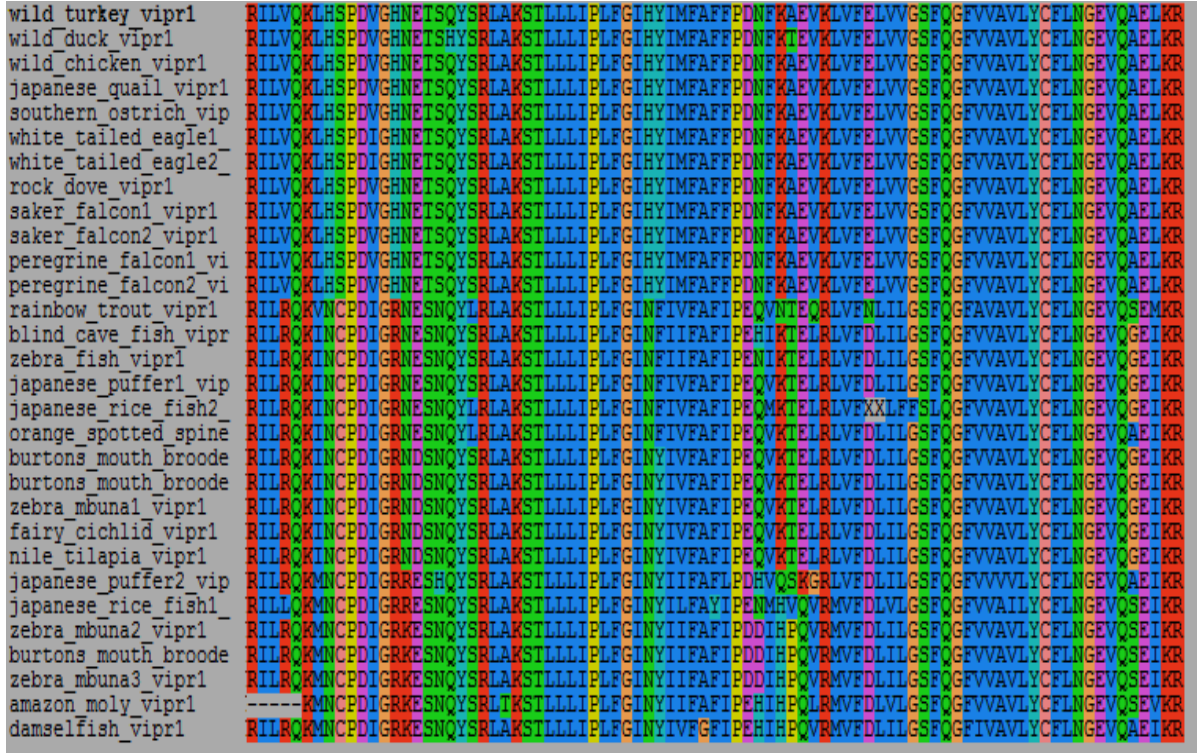
Figure 4.3: Multiple sequence alignment of vasoactive intestinal peptide receptor 1 (VIPR1) protein. The alignment shows the variations between the sequences.

4.2.4 Lineage selection in vasoactive intestinal peptide receptor 1 (VIPR1)

The phylogenetic tree for VIPR1 (shown in Figure 4.4) was constructed using FastME which is distance-based with 1000 bootstraps replicates to test the confidence of the topology. To test whether there was variable selective pressures acting on the lineages, the coding sequences for the homologs were fitted to the one ratio model (Model0) and the free ratio model (Model1) (Yang, 1998). For VIPR1, model1 fitted significantly better than model0 with P<0.05 when compared to X^2 distribution as seen in Table 4.1. This was an indication of heterogeneity in the selective pressures along the lineages.

Purifying selection was found to be acting on all lineages in VIPR1. All the lineages had

a dN/dS<1.
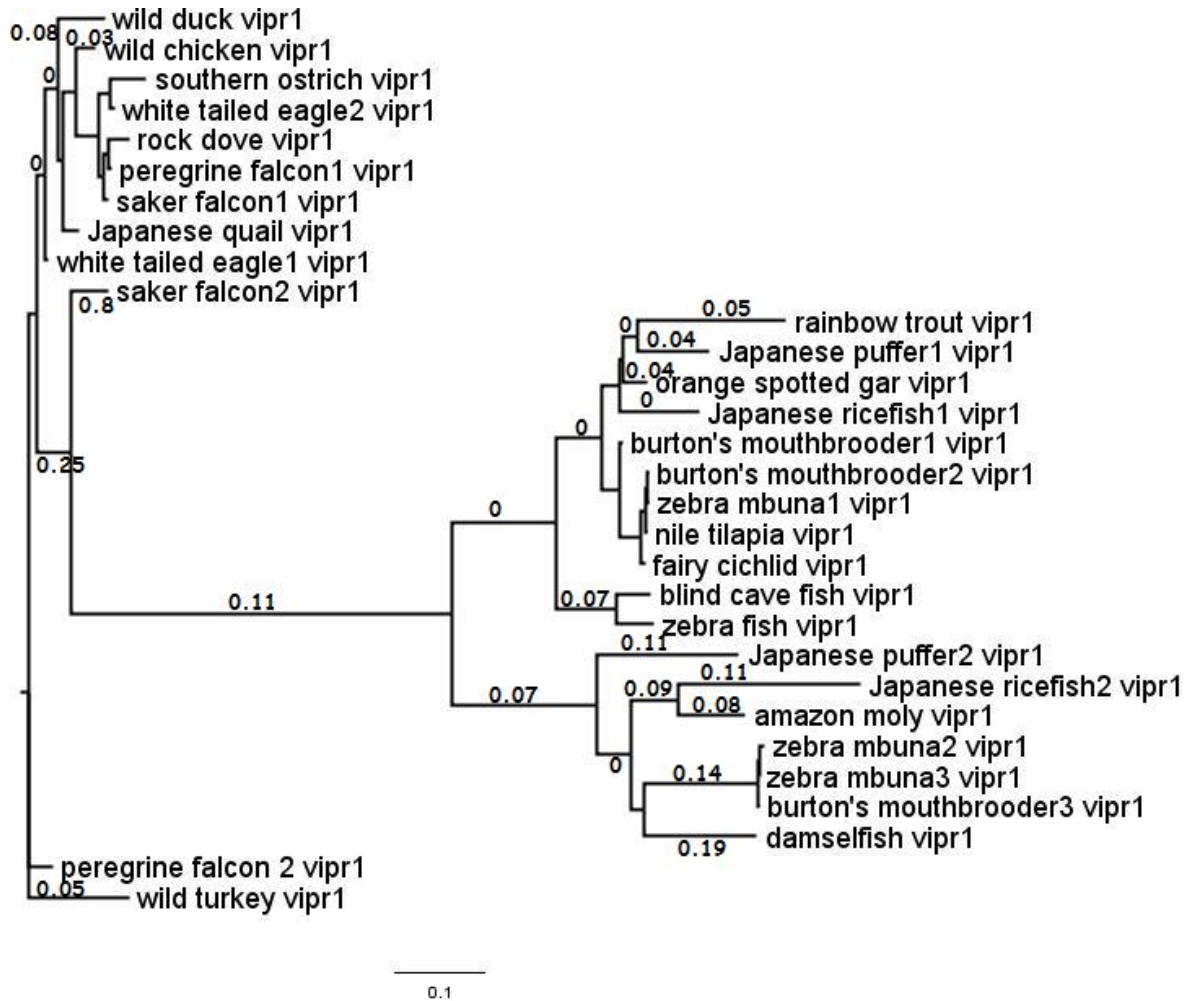
Figure 4.4: A phylogram showing evolution of vasoactive intestinal peptide receptor 1 (VIPR1) gene. The branch lengths were estimated by maximum likelihood under the free ratio model (Model 1) which assumes an independent ω for each branch. The branch lengths are drawn in proportion to the expected numbers of nucleotide substitutions per codon.

4.2.5 Lineage selection in prolactin gene

The phylogram for prolactin shown in Figure 4.5 was constructed using FastME which is distance-based with 1000 bootstraps replicates to test the confidence of the topology.

To test whether there was variable selective pressures acting on the lineages, the coding sequences for the homologs were fitted to the one ratio model (Model0) and the free ratio model (Model1) (Yang, 1998)**.** For prolactin, model1 fitted significantly better than model0 with P<0.05 when compared to X^2 distribution as seen in Table 4.1. This is an indication of heterogeneity in selective pressures along the lineages of the phylogeny below. The two branches leading to poultry were under positive selection. The branch leading to the ancestor of the wild chicken, quails and common peafowl had a high dN/dS ratio of 3.64. The rest of the lineages are under purifying selection.
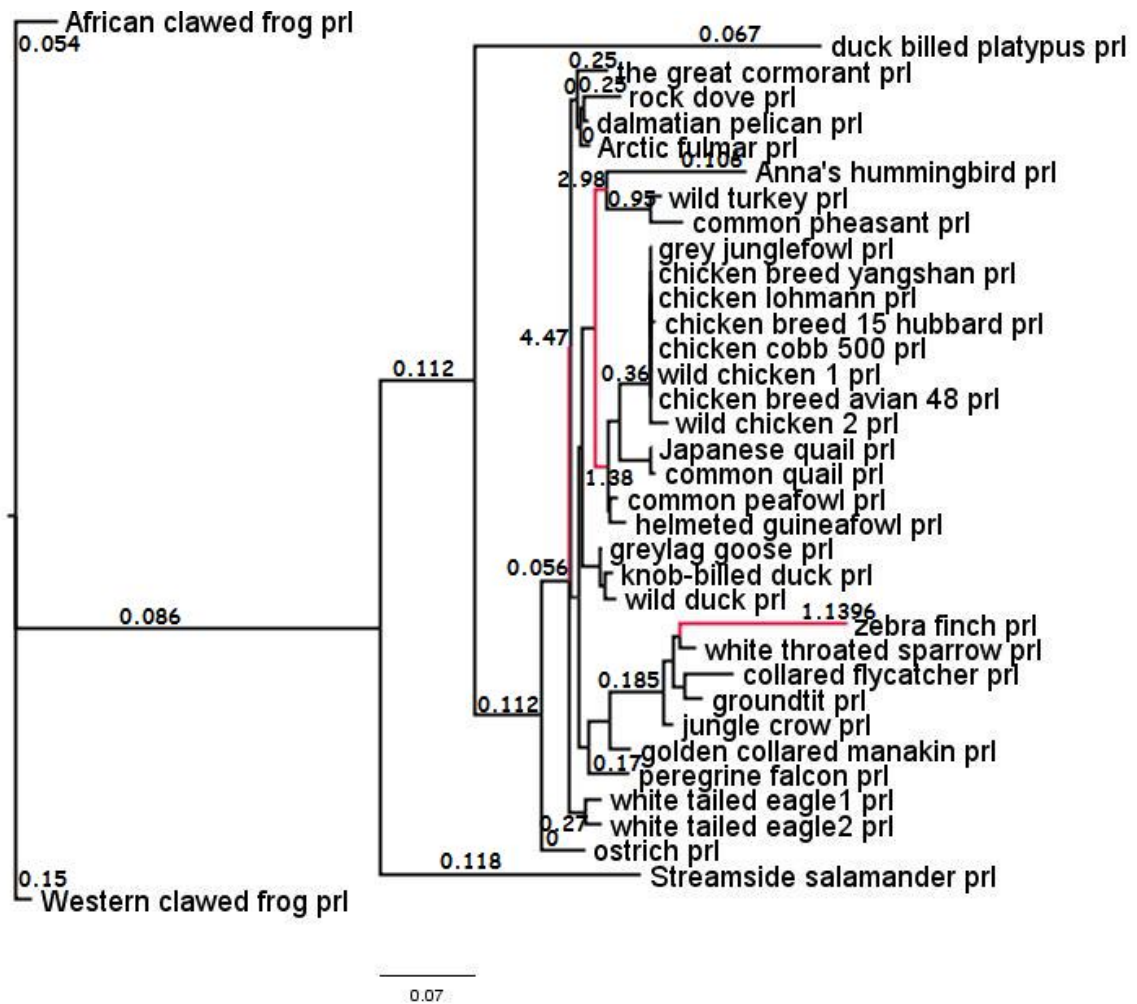
Figure 4.5: Phylogeny showing evolution of prolactin (PRL) gene. The branch lengths were estimated by maximum likelihood under the free ratio model (Model 1) which assumes an independent ω for each branch. The branch lengths are drawn in proportion to the expected numbers of nucleotide substitutions per codon. The colored branches have dN/dS>1 hence indicates positive selection.

4.3 Analysis of signatures of selection at candidate genes for growth in poultry

4.3.1 Multiple sequence alignment of growth hormone (GH)

On performing Multiple sequence alignment of the growth hormone homologues using Muscle software, the alignment in Figure 4.6 below was produced.

Among the poultry species used in this alignment: wild chicken, common quail, wild duck, greylag goose, and Southern ostrich, the Southern ostrich has the most substitutions. At amino acid position 13, alanine was substituted by threonine in common quail, wild duck and greylag goose. At position 20, proline was substituted by glutamine in the greylag goose, wild duck and Southern ostrich. At position 69, tyrosine was substituted by histidine in greylag goose, wild duck and southern ostrich. At position 70, alanine substituted threonine in Southern ostrich, at position 86 valine substituted alanine, at position 87, phenylalanine substituted proline, at position 136, tyrosine substituted phenylalanine.
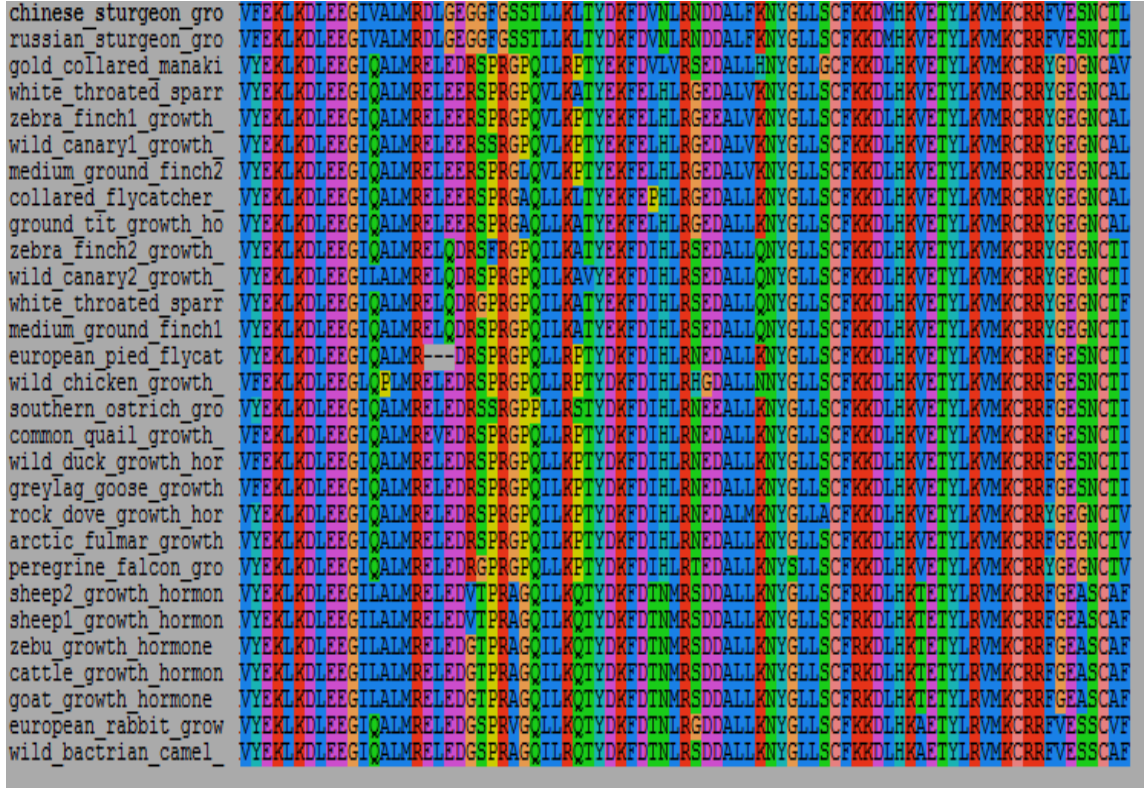
Figure 4.6: Multiple sequence alignment of growth hormone (GH) protein. The alignment shows the high level of identity between the poultry sequences. The Southern ostrich has more substitutions compared to the other poultry sequences.

4.3.2 Multiple sequence alignment of growth hormone receptor (GHR)

The alignment below (Figure 4.7) was produced by the Multiple Sequence Alignment software Muscle.

The sequences of greylag goose, knob-billed duck, and wild duck are highly similar with very few varying sites. At position 224, serine substituted alanine in these three species, at position 230, arginine substituted glutamine. At position 308, threonine substituted isoleucine in knob-billed duck, at position 337, glycine substituted glutamic acid in wild

duck. At position 595, the three species have cysteine substituting tyrosine, at position 633, aspartate substituted histidine, at position 639, the greylag goose had valine being substituted by histidine while in knob-billed duck and wild duck, aspartate substituted valine, at position 641, serine substituted alanine in the three species, at position 650, alanine substituted threonine in the three species. At position 671 of the greylag goose, alanine substituted proline while leucine substituted proline in the knob-billed at the same position, at position 676 of the knob-billed duck, arginine substituted threonine while in the wild duck, serine substituted threonine.
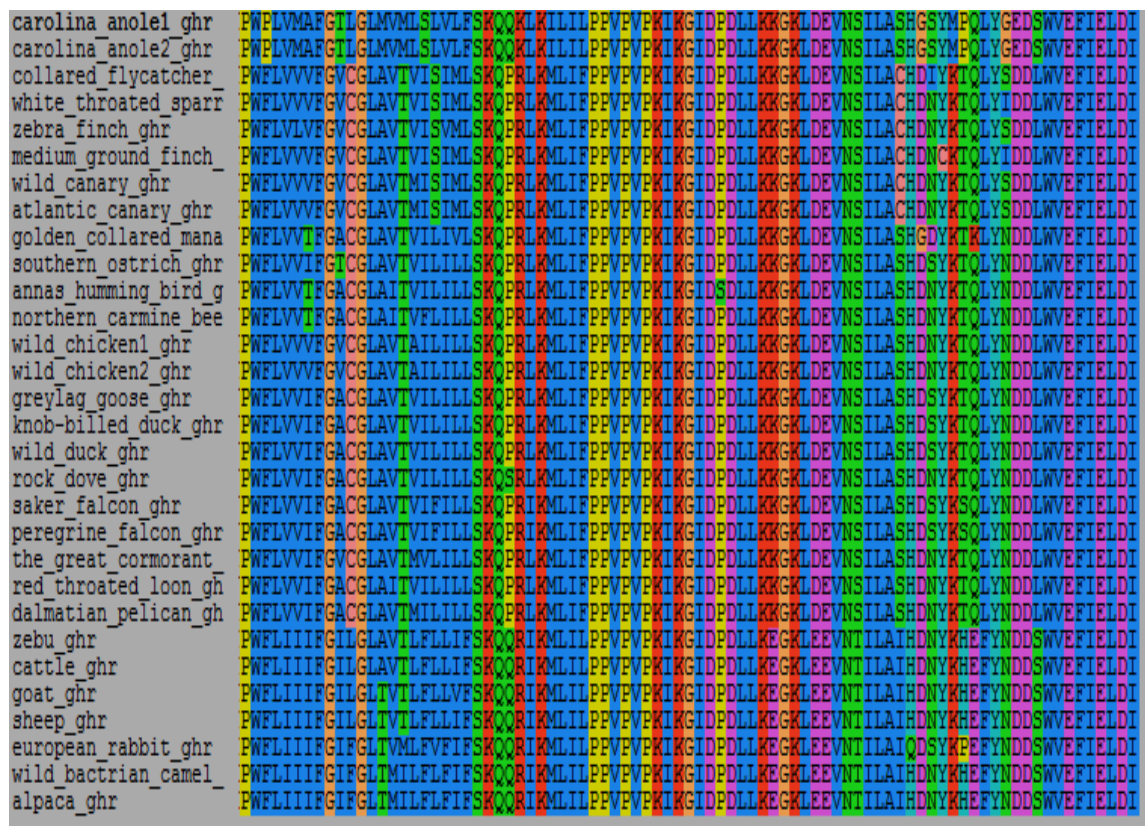
Figure 4.7: Multiple sequence alignment of growth hormone receptor (GHR) protein. The alignment shows that there are very few sites with variations between the poultry sequences.

4.3.3 Multiple sequence alignment of insulin-like growth factor 1 (IGF1) gene

The alignment below (Figure 4.8) was produced on aligning the Insulin-like growth factor 1 homologous sequences using Muscle software.

The IGF1 gene is conserved with a few substitutions. At position 3 of the wild duck, lysine was substituted by glutamate, at position 58 of the wild turkey, alanine substituted glycine, at position 66 of Yunnan Daweishan chicken breed, serine was substituted by isoleucine, at position 117 of chicken haplotype h-12 and chicken haplotype h-5, proline substituted glutamine. In addition, at position 138 of chicken haplotype h-12 and chicken haplotype h-5, asparagine substituted lysine, at position 144 of chicken haplotype h-12 and chicken haplotype h-5, leucine substituted arginine, at position 145 of wild turkey and Southern ostrich, glycine was substituted by alanine.
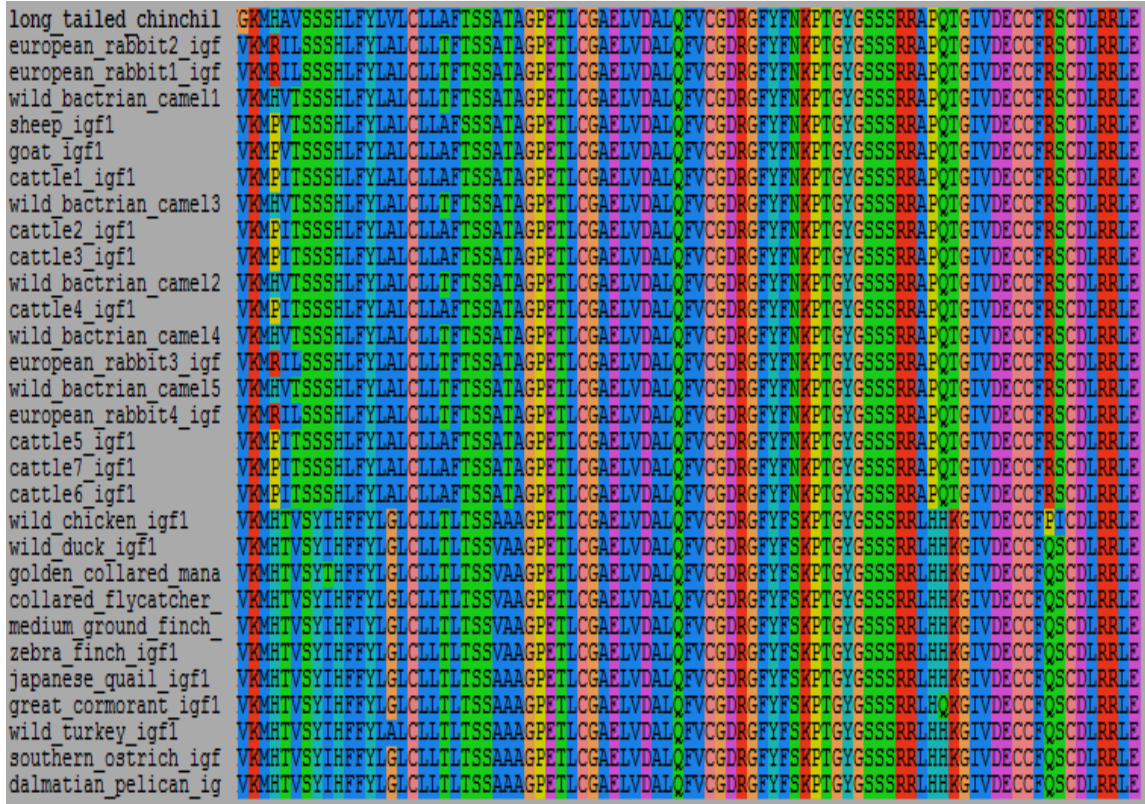
Figure 4.8: Multiple sequence alignment of insulin-like growth factor 1 (IGF1) protein. The alignment shows that this gene is highly conserved.

4.3.4 Multiple sequence alignment of insulin-like growth factor 1 receptor (IGF1R) gene

On using Muscle software to perform a Multiple sequence alignment of IGF1R, the alignment in Figure 4.9 was obtained.

The sequences of wild duck and Southern ostrich are similar and had more substitutions compared to other poultry species: wild turkey, wild chicken and Japanese quail. At position 213 of wild turkey and the Southern ostrich, valine substituted isoleucine, at position 471 of the Southern ostrich, isoleucine substituted threonine, at position 745 of

the Southern ostrich and wild duck asparagine substituted aspartate. At position 944 of wild duck and the Southern ostrich, threonine substituted alanine, at position 1117 of the Southern ostrich and wild duck, glycine substituted aspartate. At position 1291 of wild duck, threonine substituted alanine, at position 1331 of wild turkey, lysine substituted glutamic acid, at position 1332 of wild turkey, methionine substituted aspartate. At position 1474 of Japanese quail, alanine substituted proline.
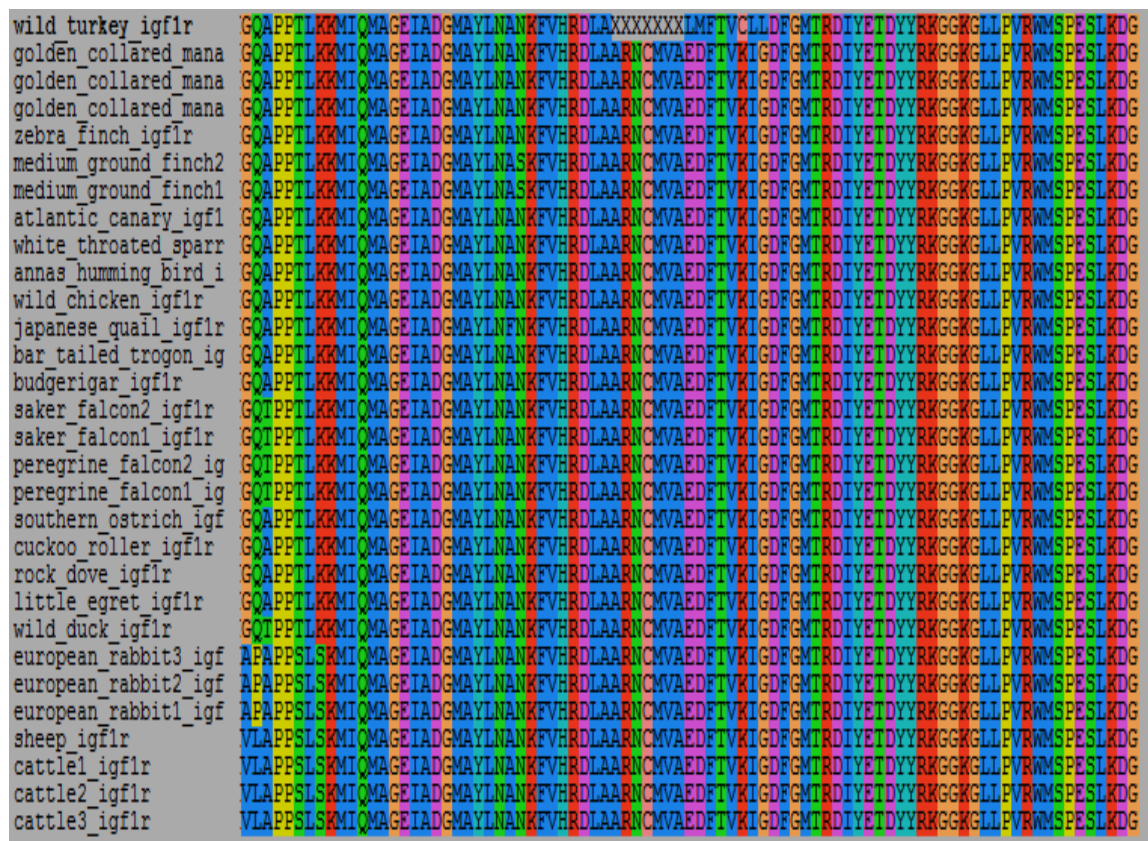
Figure 4.9: Multiple sequence alignment of insulin-like growth factor 1 receptor (IGF1R) protein. The alignment shows that the sequences of the Southern ostrich and wild duck are highly similar but differ from the other sequences.

Table 4.1: Likelihood ratio tests (LRT) to detect selection in lineages and amino acid sites. Likelihood ratio tests were done to determine fitness of a model. Probability values of <0.05 shows that a model is significant.

Gene	Model Analysis	2Δ1 Lineage= 2(InlM0-InlM1) Site= 2(InlM7-M8)	X^2 (Chi squared) value	Degrees of freedom	Probability Value
Prolactin	Lineage Analysis	2(-5519.413936+5449.921133) =138.9866	138.9866	139-71=68	<0.05
	Site Analysis	2(-5469.7794-5466.9746) =5.6096	5.6096	74-72=2	>0.05
VIP1	Lineage Analysis	2(-529.499090+514.296366) =30.41	30.41	115-59=56	>0.05
	Site Analysis	2(-517.381707+517.979078) =1.194742	1.19	62-60=2	>0.05
VIPR1	Lineage Analysis	2(-9226.699685+9164.329783) =124.7	124.7	115-59=56	<0.05
	Site Analysis	2(-9013.544+9005.761) =15.566	15.566	62-60=2	<0.05
GH	Lineage Analysis	2(-4704.661507+4661.203348) =86.9	86.9	111-57=54	<0.05
	Site Analysis	2(-4634.798616+4634.799225) =0.001218	0.001218	60-58=2	>0.05
GHR	Lineage Analysis	2(-11526.92548+11486.241214) =81.568540	81.568540	127-65=62	<0.05
	Site Analysis	2(-11295.23339+11293.456950) =3.552886	3.552886	68-66=2	>0.05
IGF1	Lineage Analysis	2(-2076.081066+2032.048315) =88.065502	88.065502	151-77=74	>0.05
	Site Analysis	2(-2017.692136+2015.465467) =4.853338	4.853338	80-78=2	>0.05
IGF1R	Lineage Analysis	2(-12701.06843+12641.506653) =119.1	119.1	115-57=56	<0.05

4.3.5 Lineage selection in growth hormone (GH) gene

The phylogram of GH shown in Figure 4.10 below was constructed using distance-based FastME with 1000 bootstrap replicates to test the confidence of the topology. To test whether there was variable selective pressures acting on the lineages, the coding sequences for the homologs were fitted to the one ratio model (model0) and the free ratio model (model1) (Yang, 1998). For GH, model1 fitted significantly better than model0 with P<0.05 when compared to X^2 distribution as seen in Table 1. This is an indication of heterogeneity in selective pressures along the lineages as shown in Figure 4.10.

In GH, positive selection was detected in the branch leading to cattle, sheep, goat and camel (shown in red). The dN/dS was 4.98 which is greater than 1. In the rest of the lineages, GH was under purifying selection as indicated by dN/dS of <1.
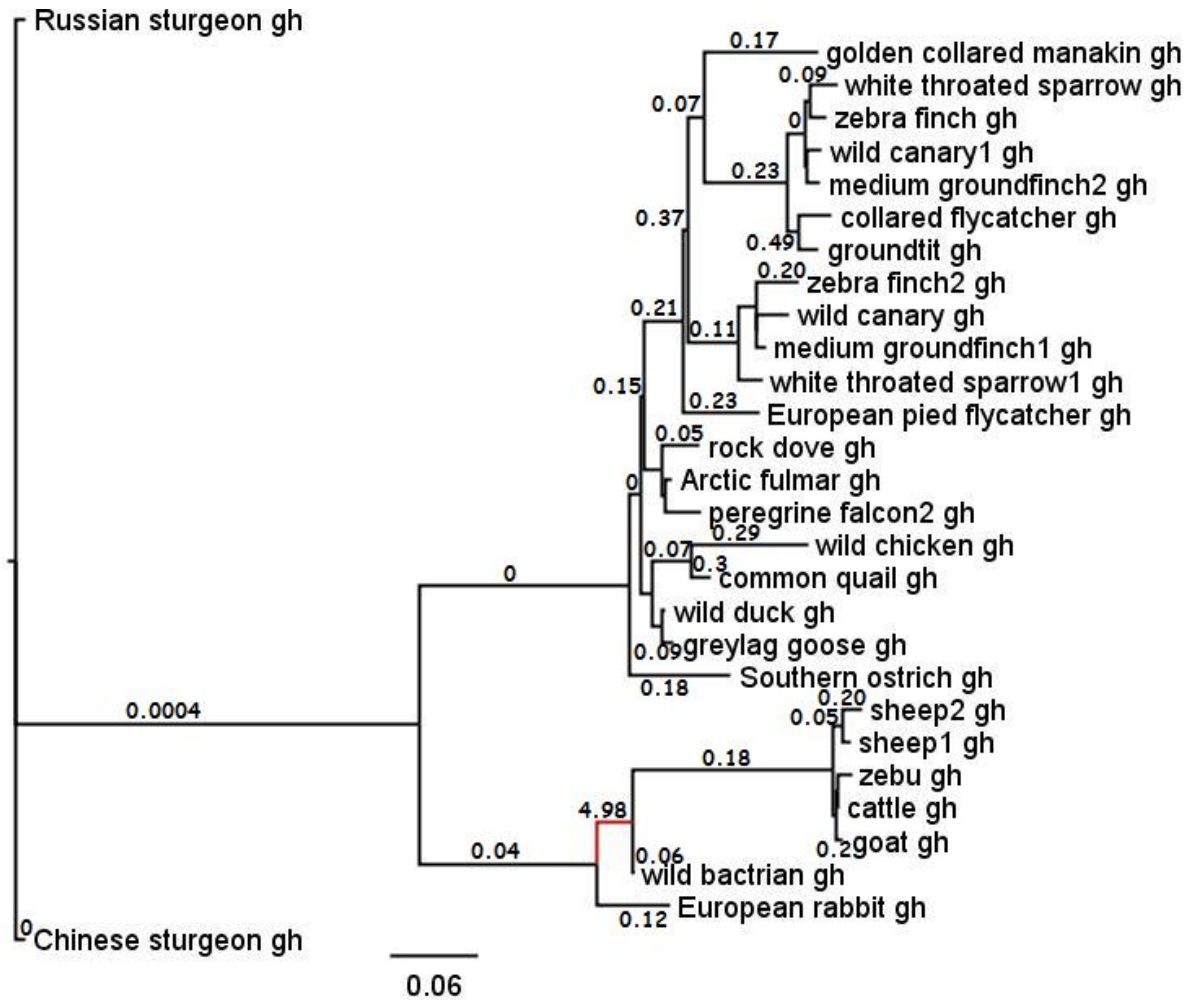
Figure 4.10: Phylogenetic tree showing evolution of growth hormone (GH) gene.

The branch lengths are in proportion to the number of nucleotide substitutions per codon. There is

heterogeneity in the dN/dS values in the branches. dN/dS>1 indicate positive selection (shown in red).

4.3.6 Lineage selection in insulin-like growth factor 1 receptor (IGF1R) gene

The phylogram of IGF1R (Figure 4.11) was constructed using distance-based FastME

with 1000 bootstrap replicates to test the confidence of the topology.

To test whether there was variable selective pressures acting on the lineages, the coding sequences for the homologs were fitted to the one ratio model (model0) and the free ratio model (model1) (Yang, 1998). In the Lineages, P values of < 0.05 confirm heterogeneity in selection pressures. For IGF1R, model1 fitted significantly better than model0 with $P<0.05$ when compared to X^2 distribution as seen in Table 1. This is an indication of heterogeneity in selective pressures along the lineages as shown in Figure 4.11.

Purifying selection was found to be acting on all lineages in IGF1R. All the lineages had a dN/dS<1.
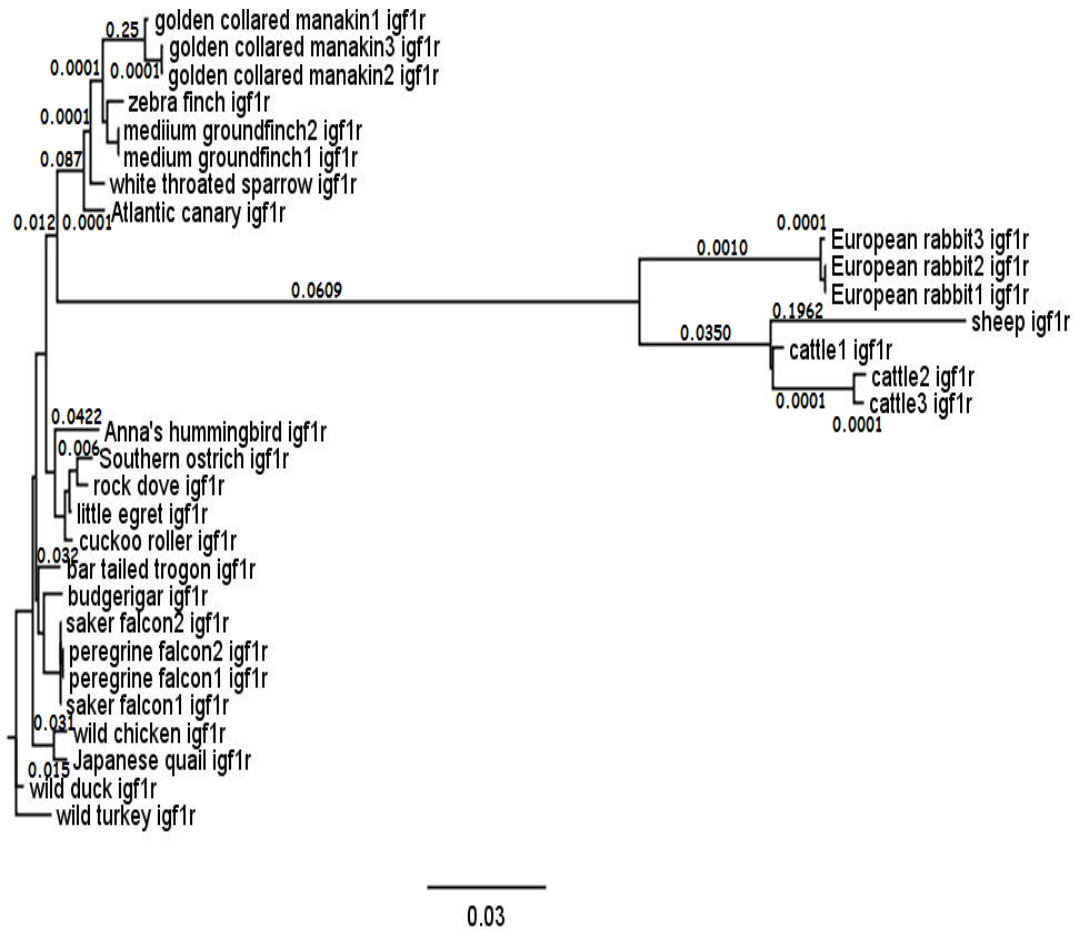
Figure 4.11: Phylogenetic tree of the insulin-like growth factor 1 receptor (IGF1R) gene. The tree was constructed using FastME with 1000 bootstraps. The branch lengths were estimated using maximum likelihood under the free ratio model (M1) which assumes an independent ω for each branch. The branch lengths are drawn in proportion to the expected numbers of nucleotide substitutions per codon.

4.3.7 Lineage selection in growth hormone receptor (GHR) gene

The phylogram of GHR shown in Figure 4.12 below was constructed using distance-based FastME with 1000 bootstrap replicates to test the confidence of the topology.

52

To test whether there was variable selective pressures acting on the lineages, the coding sequences for the homologs were fitted to the one ratio model (model0) and the free ratio model (model1) (Yang, 1998). For GHR, model1 fitted significantly better than model0 with P<0.05 when compared to X^2 distribution as seen in Table 1. This is an indication of heterogeneity in selective pressures along the lineages as shown in Figure 4.12.

Purifying selection was found to be acting on all lineages in GHR. All the lineages had a dN/dS<1.
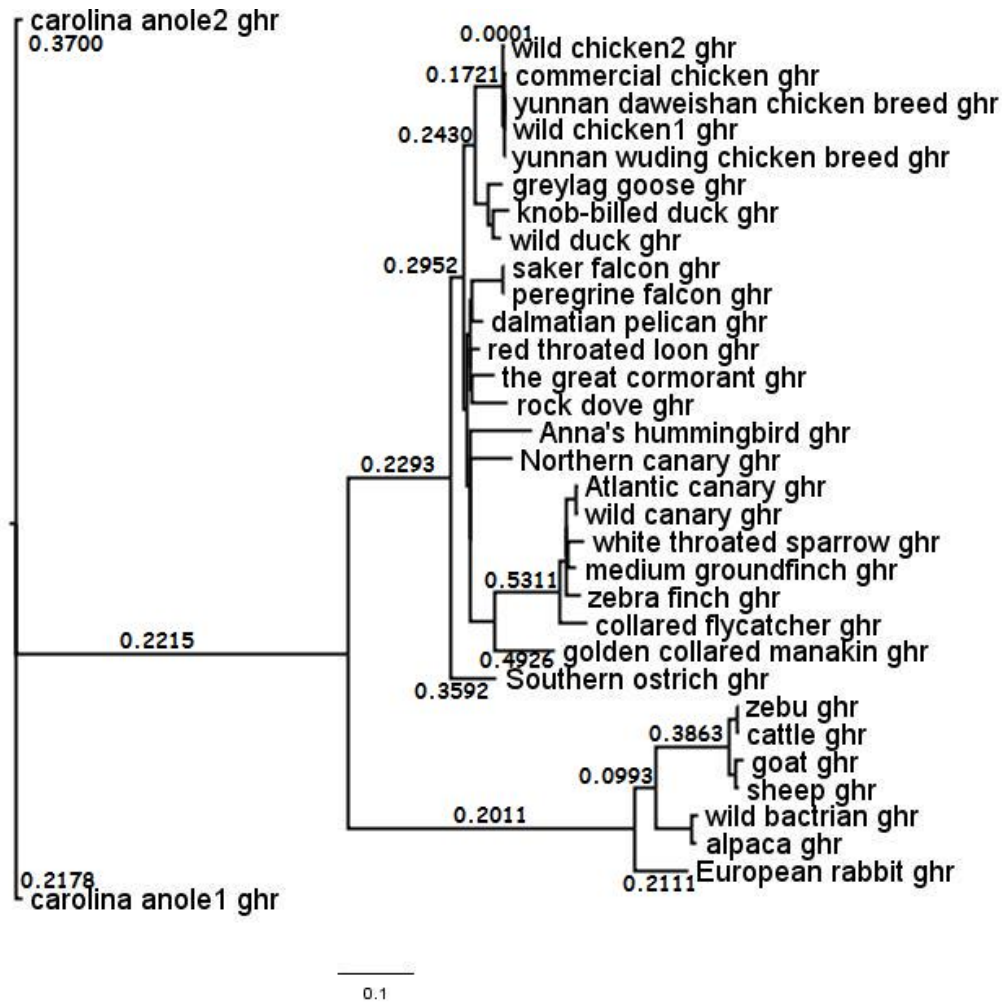
Figure 4.12: Phylogenetic tree showing evolution of growth hormone receptor (GHR) gene. The branch lengths were estimated using maximum likelihood under the free ratio model (M1) which assumes an independent ω for each branch. The branch lengths are drawn in proportion to the expected numbers of nucleotide substitutions per codon.

4.3.8 Amino acid site selection

To test the hypothesis of different selective pressure on the amino acid sites, twice the log likelihood difference of Model7 and Model8 was done for all the proteins as shown in Table 4.1.

The Multiple sequence alignment of insulin-like growth factor 1 receptor (IGF1R) with Isoleucine under positive selection is shown in Figure 4.13. The Likelihood ratio test showed that in GH and IGF1R genes, there are different selective pressures acting on each amino acid site. Posterior probability analysis identified positive selection on amino acid Isoleucine at site 460 of the IGF1R to be under positive selection as it had a posterior probability of 0.985. This was greater than the predetermined cut-off value of >0.95 for inferring that a site belongs to a class which has a dN/dS of > 1.

Figure 4.13: Multiple sequence alignment of insulin-like growth factor 1 receptor (IGF1R) with site under positive selection. Posterior probability analysis identified isoleucine at position 460 to be under positive selection. The site is shown in red.

The raptor X server predicted the 3D structure of insulin-like growth factor 1 receptor (IGF1R) which is shown in Figure 4.14. The amino acid site under positive selection

was located on the 3D structure. Pymol was used to visualize the structure with the site under selection.

(green)	Receptor L Domain
(olive)	Furin-like cysteine rich domain
(blue)	Protein tyrosine kinase domain
(cyan)	Fibronectin type III domain
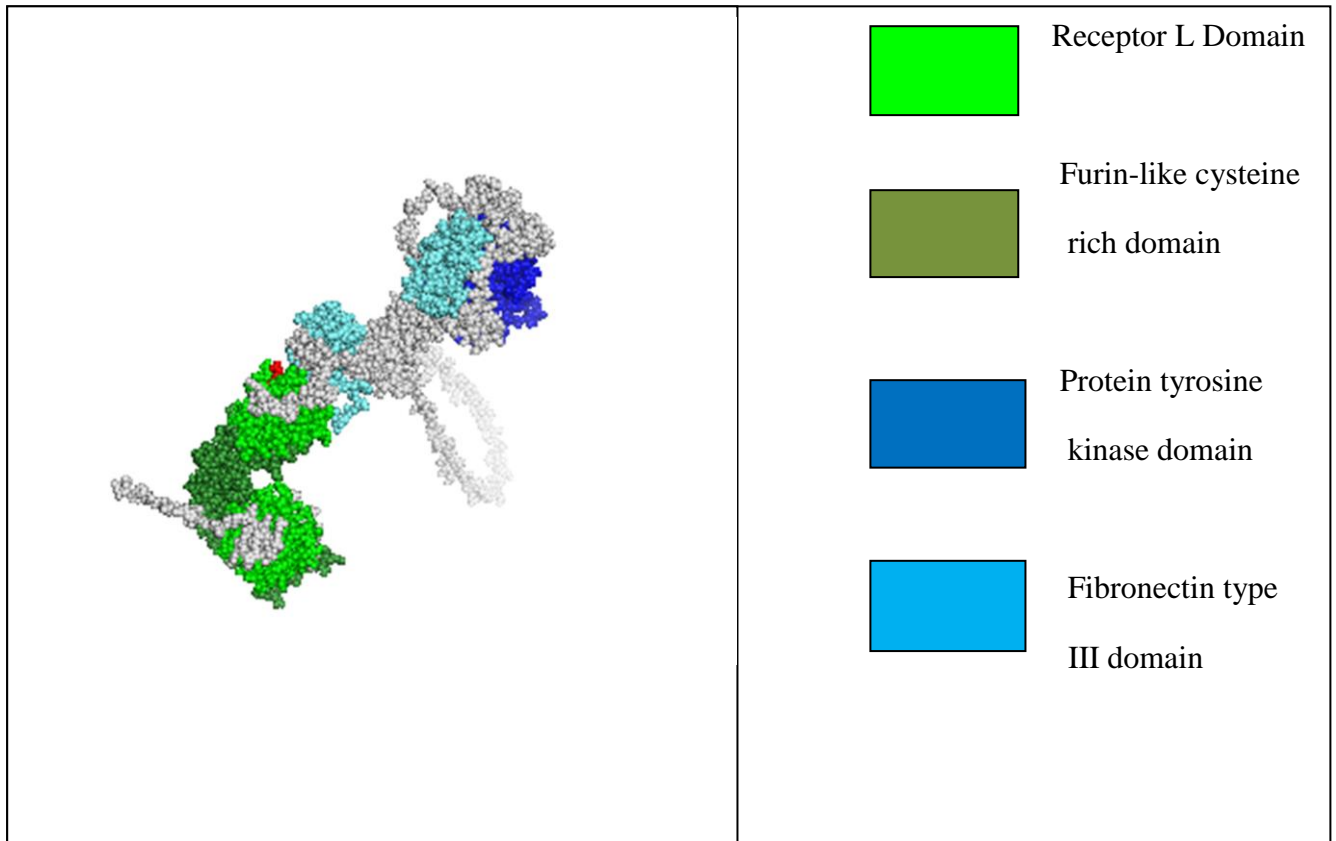
Figure 4.14: The 3D structure of insulin-like growth factor 1 receptor showing the different domains. The site shown in red is predicted to be under positive selection by Model 8. The site is located in the Receptor L domain which is important for binding insulin-like growth factor 1.

4.4 Protein 3D structure prediction for genes implicated in egg production

4.4.1 Predicted 3D structures of prolactin protein

In Table 4.2 below, the secondary and tertiary structures of prolactin from different poultry species were predicted. There was a small variation in the secondary structure but the tertiary structure was similar.

Table 4.2: Predicted 3D structures of prolactin protein. The table below shows the secondary and tertiary structures of prolactin from different poultry species.

Prolactin protein structures	Helices	Beta sheets	Loops	Domain interval	Solvent accessibility
Common peafowl	62%	0%	37%	31-229	50%
Common pheasant	63%	0%	35%	31-229	50%
Common quail	60%	0%	39%	31-229	50%
Helmeted guinea fowl	65%	0%	34%	31-229	50%
Japanese quail	58%	0%	41%	31-229	50%
Knob-billed duck	65%	0%	34%	31-229	50%
Ostrich	65%	0%	34%	31-229	50%
Wild chicken	63%	0%	36%	31-229	50%
Wild duck	59%	0%	40%	31-229	50%
Wild turkey	62%	0%	37%	31-229	50%

4.4.2 Predicted 3D structures vasoactive intestinal peptide 1 protein

In Table 4.3 below, the secondary and tertiary structures of vasoactive intestinal peptide 1 from different poultry species were predicted. There was some variation in the secondary structure but the tertiary structure was similar.

Table 4.3: Predicted protein 3D structures of vasoactive intestinal peptide 1. The secondary and tertiary structures of VIP1 from different poultry species are shown in the table below.

Vasoactive intestinal peptide 1 (VIP1) protein structures	Helices	Beta sheets	Loops	Domain interval	Solvent accessibility
Wild chicken 1	48%	0%	51%	89-115, 129-155	50%
Wild chicken 2	33%	0%	66%	89-115, 129-155	50%

4.4.3 Predicted 3D structures of vasoactive intestinal peptide receptor 1 protein

In Table 4.4 below, the secondary and tertiary structures of vasoactive intestinal peptide receptor 1 from different poultry species were predicted. The secondary structure was similar but the tertiary structure varied between the species.

58

Table 4.4: Predicted protein 3D structures of vasoactive intestinal peptide receptor 1. The secondary and tertiary structures of VIPR1 from different poultry species are shown in the table below.

Vasoactive intestinal peptide receptor 1 (VIPR1) protein structures	Helices	Beta sheets	Loops	Domain interval	Solvent accessibility
Japanese quail	52%	9%	38%	50-114, 127-373	50%
Wild chicken	52%	9%	38%	22-86, 99-345	50%

4.5 Protein 3D structure prediction for genes implicated in growth

4.5.1 Predicted 3D structures of growth hormone protein

In Table 4.5 below, the secondary and tertiary structures of growth hormone from different poultry species were predicted. The secondary structure was similar with minimum variation in the tertiary structure.

Table 4.5: Predicted protein 3D structures of growth hormone. The secondary and

tertiary structures of GH from different poultry species are shown in the table below.

Growth hormone (GH) protein structures	Helices	Beta sheets	Loops	Domain interval	Solvent accessibility
Common quail	56%	0%	43%	35-214	50%
Greylag goose	56%	0%	43%	36-214	50%
Wild chicken	55%	0%	43%	35-214	50%
Wild duck	56%	0%	43%	35-214	50%

4.5.2 Predicted 3D structures of growth hormone receptor protein

In Table 4.6 below, the secondary and tertiary structures of growth hormone receptor from different poultry species were predicted. There were notable variations in the secondary and tertiary structures in the different species.

Table 4.6: Predicted protein 3D structure of growth hormone receptor. The secondary

and tertiary structures of GHR from different poultry species is shown in the table above.

Growth hormone receptor (GHR) protein structures	Helices	Beta sheets	Loops	Domain interval	Solvent accessibility
Greylag goose	6%	19%	73%	29-108, 123-223, 292-590	50%
Southern ostrich	9%	19%	70%	24-233	50%
Wild chicken 1	9%	21%	69%	24-230	50%
Wild chicken 2	10%	21%	68%	24-230	50%
Wild duck	9%	19%	17%	24-233	50%

4.5.3 Predicted 3D structures of insulin-like growth factor 1 protein

In Table 4.7 below, the secondary and tertiary structures of insulin-like growth factor 1

from different poultry species were predicted. There were variations in the secondary

structure but the tertiary structure was similar in all the species.

Table 4.7: Predicted protein 3D structure of insulin-like growth factor 1. The

secondary and tertiary structures of IGF1 from different poultry species are shown in the table below.

Insulin-like growth factor 1 (IGF1) protein structures	Helices	Beta sheets	Loops	Domain interval	Solvent accessibility
Japanese quail	32%	5%	62%	49-116	50%
Southern ostrich	37%	9%	53%	49-116	50%
Wild chicken	29%	5%	64%	49-114	50%
Wild duck	32%	5%	62%	49-116	50%
Wild turkey	31%	5%	62%	49-116	50%

4.5.4 Predicted 3D structures of insulin-like growth factor 1 receptor protein

In Table 4.8 below, the secondary and tertiary structures of insulin-like growth factor 1 receptor from different poultry species were predicted. There was minimum variation in the secondary structure but notable differences in the tertiary structure of the different species.

Table 4.8: Predicted protein 3D structure of insulin-like growth factor 1 receptor.

The secondary and tertiary structures of IGF1R from different poultry species is shown in the table below.

Insulin-like growth factor 1 receptor (IGF1R) protein structures	Helices	Beta sheets	Loops	Domain interval	Solvent accessibility
Japanese quail	16%	24%	59%	51-161, 75-330, 230-277, 351-464, 495-605, 610-653, 835-924, 993-1269, 1000-1267	50%
Southern ostrich	14%	27%	58%	86-672, 673-989, 1041-1350	50%
Wild chicken	17%	23%	58%	86-672, 673-989, 932-972, 1041-1350	50%
Wild duck	16%	24%	59%	18-606, 606-922, 932-972, 973-1283	50%

4.6 Predicted 3D structure comparison of proteins implicated in egg production

4.6.1 Comparison of predicted 3D prolactin structures

The DeepAlign server was used to obtain the 3D structure alignment of the prolactin structures (Figure 4.15) using their sequences.

The 3D structural comparison of common peafowl, common pheasant, common quail, helmeted guinea fowl, Japanese quail, knob-billed, ostrich, wild chicken, wild duck prolactin produced a template modeling (TM) score of 0.901. This showed that these structures were highly similar and they were likely to share a similar fold.
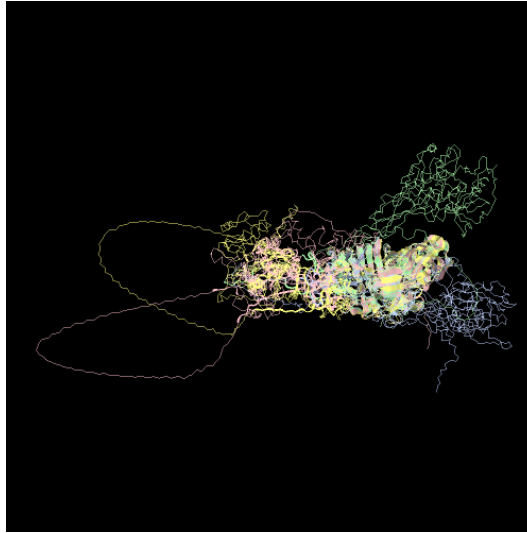
Lali: 117 RMSD: 0.79 uGDT(GDT): 116(81) TMscore: 0.901

Figure 4.15: Predicted 3D structure comparison of peafowl, common pheasant, common quail, helmeted guinea fowl, Japanese quail, knob-billed, ostrich, wild chicken, and wild duck prolactin. The scores represent 3D similarity scores. Template modeling (TM) score values of >0.5 meant that the structures were similar.

4.6.2 Comparison of predicted 3D vasoactive intestinal peptide 1 structures

On sending the sequences of the vasoactive intestinal peptide 1 (VIP1) sequences to the DeepAlign server, a structural alignment of the structures was obtained as shown in Figure 4.16 below.

 The 3D structures of wild chicken 1 and wild chicken 2 VIP1 were not similar and therefore do not share a similar fold as their structural alignment gave a template modeling (TM) score of 0.199.
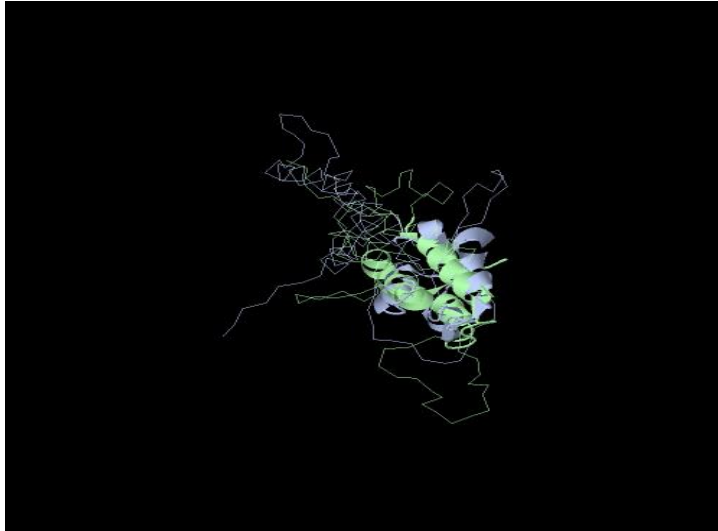
Lali: 57 RMSD: 5.19 UGDT(GDT): 26(16) TMscore: 0.199

Figure 4.16: Predicted 3D structure comparison of wild chicken 1 and wild chicken 2 vasoactive intestinal peptide 1. The 3D similarity scores were used to evaluate similarity of scores. Template modeling (TM) score of <0.5 meant that the structures were not similar.

4.6.3 Comparison of predicted 3D vasoactive intestinal peptide receptor 1 structures

The DeepAlign server produced the structural alignment of the vasoactive intestinal peptide receptor 1 (VIPR1) structures shown in Figure 4.17 below.

From the 3D structural comparison of Japanese quail and wild chicken VIPR1, it was evident that these structures were highly similar and shared a fold as they had a template modeling (TM) score of 0.65.
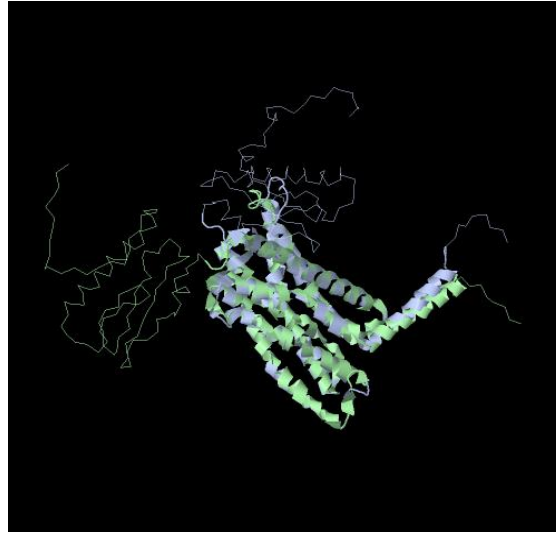
Lali: 288 RMSD: 1.96 uGDT(GDT): 270(66) TMscore: 0.675

Figure 4.17: Predicted 3D structure comparison of Japanese quail and wild chicken vasoactive intestinal peptide receptor 1. The similarity scores shown were for evaluating similarity. Template modeling (TM) score of >0.5 meant that the structures were similar.

4.7 Predicted 3D structure comparison of proteins implicated in growth

4.7.1 Comparison of predicted 3D growth hormone structures

The different sequences of growth hormone (GH) were sent to DeepAlign for structural comparison. Figure 4.18 below shows the structural superimposition.

The template modeling (TM) score of the 3D structural comparison of common quail, greylag goose, wild chicken, and wild duck GH was 0.919. These structures were very likely to share a fold and they were similar.
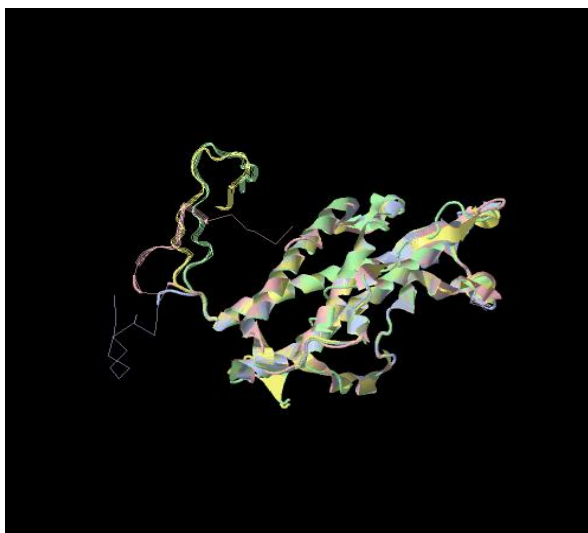
Lali: 189 RMSD: 1.13 uGDT(GDT): 185(91) TMscore: 0.919

Figure 4.18: Predicted 3D structure comparison of common quail, greylag goose, wild chicken, and wild duck growth hormone. The scores represent 3D similarity scores. The scores were used in evaluating similarity. Template modeling (TM) score of >0.5 meant that the structures were similar.

4.7.2 Comparison of predicted 3D growth hormone receptor structures

DeepAlign was used to compare the different growth hormone receptor (GHR) structures. Figure 4.19 below shows the structural alignment of the structures. On the alignment of the 3D structures of greylag goose, southern ostrich, wild chicken and wild duck GHR, the template modeling (TM) score was 0.880. This evidently showed that these structures were highly similar and they were likely to share a fold.
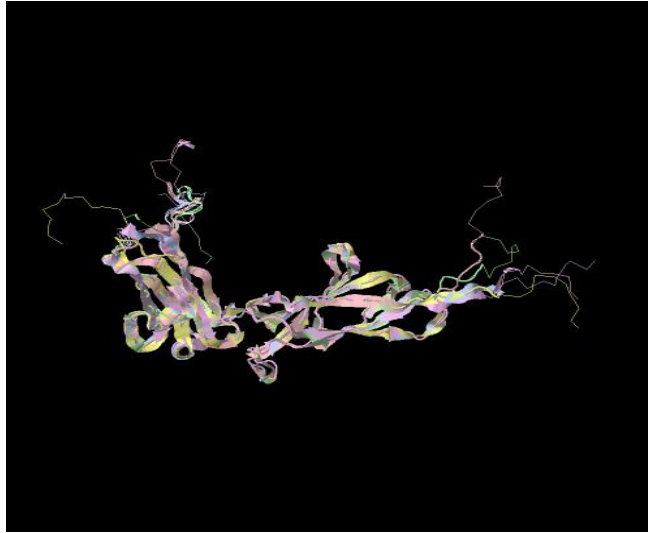
Lali: 205 RMSD: 1.36 uGDT(GDT): 198(87) TMscore: 0.880

Figure 4.19: Predicted 3D structure comparison of greylag goose, Southern ostrich, wild chicken, and wild duck growth hormone receptor. The scores were used in evaluating 3D similarity. The structures were similar as they had a template modeling (TM) score of > 0.5.

4.7.3 Comparison of predicted 3D insulin-like growth factor 1 structures

The sequences of the different insulin-like growth factor 1 (IGF1) structures were sent to DeepAlign server to obtain the structural alignment shown below (Figure 4.20). The 3D structural comparison of Japanese quail, southern ostrich, wild chicken, and wild duck IGF1 had a template modeling (TM) score of 0.615. These structures were thus similar and they were likely to share a fold.
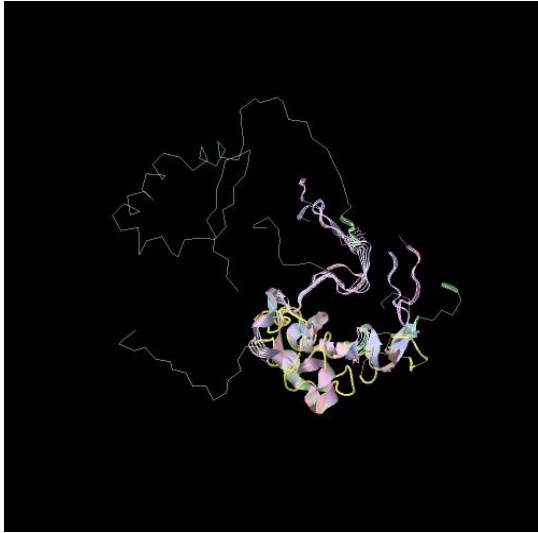
Lali: 61 RMSD: 3.35 uGDT(GDT): 43(44) TMscore:0.615

Figure 4.20: Predicted 3D structure comparison of Japanese quail, southern ostrich, wild chicken, and wild duck insulin-like growth factor 1. The TMscore was used to evaluate 3D structural similarity with template modeling (TM) score of >0.5 indicating similarity of the structures.

4.7.4 Comparison of predicted 3D insulin-like growth factor 1 receptor structures

DeepAlign was used to compare the different insulin-like growth factor 1 receptor (IGF1R) structures. Figure 4.21 below shows the structural alignment of the structures. The 3D structures of Japanese quail, southern ostrich, wild chicken, and wild duck IGF1R were not similar and were therefore less likely to share a fold. This was evident from the structural alignment that gave a template modeling (TM) score of 0.481.
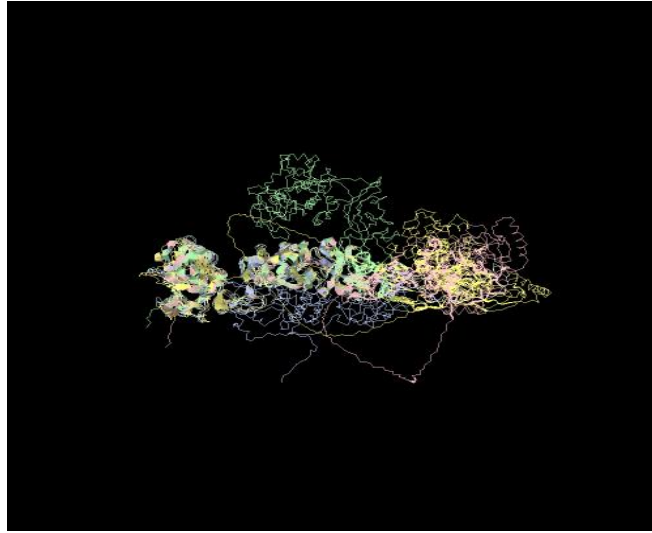
Lali: 597 RMSD: 2.77 uGDT(GDT): 522(42) TMscore: 0.481

Figure 4.21: Predicted 3D structure comparison of Japanese quail, southern ostrich, wild chicken, and wild duck insulin-like growth factor 1 receptor. The template modeling (TM) score was used to evaluate 3D structural similarity of the structures. The structures were not similar as they had a TM score of <0.5.

CHAPTER FIVE

DISCUSSION

This study supported the hypothesis that there is variation at the selected candidate genes involved in growth and the production of eggs in poultry as a result of Natural selection.

5.1 Analysis of signatures of selection in genes for egg production

5.1.1 Prolactin

In prolactin, the codon substitution models identified positive selection on the lineage leading to domesticated birds (Figure 4.5). This implies that domestication has played a role in shaping the prolactin gene in poultry. Domestication causes a rapid physical, physiological, and behavioral change (Cheng, 2010).

The domestication process is cumulative and happens over many generations (Cheng, 2010). During domestication, the animals which have been selected adapt to the new human environments by integrating the changes in their genetic makeup, psychology, physiology, and behavior (Cheng, 2010). Chickens were domesticated 5,000-10,000 years ago through the taming of a few jungle fowls (Underhill, 1997). Only 4-6 eggs are laid by the jungle fowl in a year. In contrast, a commercial laying hen lays eggs exceeding 300 in a year. This is because commercial layers have been subjected to artificial selection for higher production (Cheng, 2010). Prolactin has many and diverse roles but in birds, it is implicated in the onset and maintenance of broodiness (Wilkanowska *et al*., 2014). Broodiness causes avian egg production to be low (Usman

et al., 2014). Through domestication, the prolactin gene has been subjected to pressures to meet human demands. The other species in this study are not domesticated. This is probably the reason these species are under a different selection pressure.

The ratio of non-synonymous mutations to synonymous mutations was greater than 1 implying that on the poultry lineage there was an increase egg production.

5.1.2 Vasoactive intestinal peptide receptor 1

All lineages in vasoactive intestinal peptide receptor 1 (VIPR1) had a dN/dS<1 which was an indication that this protein is under purifying selection (Figure 4.4). This is an indication that VIPR1 has endured strong selection pressures during breeding. Zhou *et al*., (2008) also concluded that VIPR1 had endured strong selection pressures in chickens during breeding.

Broodiness is a trait which is polygenic (Romanov *et al*., 2002). The role of prolactin is initiating and maintaining incubation in birds (Zhou *et al*., 2008). At the hypothalamus, the secretion of prolactin is controlled by vasoactive intestinal peptide 1 while at the pituitary level the secretion is under the control of VIP receptor 1. Thus VIPR1 may be a candidate gene for broodiness in avian species (Zhou *et al*., 2008). In a study done by Zhou *et al*., 2008, there was variation in VIPR1 in different chicken populations. These variations were associated with broodiness. Broodiness has been a challenge to eradicate even with domestication and artificial selection (Zhou *et al*., 2008).

Purifying selection has also been detected in several protein-coding sequences in HIV-1. Many of these sequences are involved in regulating gene expression in HIV. Some of the sequences include 3' Long terminal repeat (LTR) which encode the Nef protein (Das *et al*., 1998). It also contains some regulatory elements. For example in the pol gene, there is an intragenic nuclease hypersensitive region which has a role in regulating gene expression.

Additionally, in the env gene is the rev-responsive element (RRE) which transports unspliced or RNA that partially spliced from the nucleus of the infected cell to the cytoplasm (Ngandu *et al*., 2008). These regions are important for viral viability hence are under purifying selection to preserve their functions (Ngandu *et al*., 2008).

Likewise in VIPR1 as in HIV-1, purifying selection acts to preserve the function. As a result of purifying selection, the function of VIPR1 on prolactin secretion is maintained. Although broodiness leads to a decrease in egg production, it is an important condition when one wants to use the natural process to produce chicks (FAO, 2010).

5.2 Analysis of signatures of selection in genes for growth

5.2.1 Growth hormone

In this study, the evolution of growth hormone (GH) is different in the various taxonomic groups. Positive selection was identified in the lineage leading to ruminants and artiodactyls, while purifying selection was detected in the other lineages (Figure

4.10). This is an indication that the gene has had different patterns of evolution in different species.

There are two features of interest in the molecular evolution of GH: (i) Variation in evolutionary rate (ii) Multiple duplications of the gene (Forsyth and Wallis, 2002). The sequence of GH is very conserved in mammals (Forsyth and Wallis, 2002). The differences in the sequences of mammalian species occurs at a few amino acid sites. The rate of GH evolution has been very slow in eutherians with the rate having increased 25-50 times during the evolution of primates and during the evolution of artiodactyls (Forsyth and Wallis, 2002).

The evolution of GH in mammals has been slow with most changes (~85%) having occurred during the periods of rapid change. As a result of the bursts of evolution, GH shows specie specificity. The specificity is due to single substitutions in the GH and its receptor (Forsyth and Wallis, 2002). For example GH from mammals differs from that of non-primate mammals at 35% of the residues. This is reflected in the multiple sequence alignment of GH in Figure 4.6. Further, the sequences of the different artiodactyls differ at a few amino acid sites while their sequences differ at multiple amino acid sites from other species. Hence GH from non-primates is non-functional in mammals as it binds very poorly to the receptor. The specie specificity correlates with differences in the binding of the growth hormone receptor (GHR) and the biological actions (Buggiotti and Primmer, 2006). These bursts of change occurred for a short period after which the rate fell back to the slower rate. Although there are three possible causes for the variation in the evolution rate: (a) The variability being an artifact because

of comparing non-orthologous proteins (b) The variability reflected periods of evolution of GH when the function was lost so the sequence change was caused by neutral mutations, and (c) The bursts of change were driven by positive natural selection. The third explanation was found to be the most compelling. One argument supporting the third explanation was that the episodes of rapid change occurred on the mature protein. In the mature GH protein, the rate of non-synonymous to synonymous mutation rose from 0.02 (basal) to 1.37 (artiodactyls) and 1.27 (primates) (Buggiotti and Primmer, 2006). This was evidence that the driving force was positive selection. In my study, the ratio of the non-synonymous to synonymous changes in artiodactyls was 4.98. It is evident that during this period there was accelerated rate of evolution which was driven by positive selection (Forsyth and Wallis, 2002).

The pattern of evolution in the avian growth hormone is different from mammals because of the difference in function (Buggiotti and Primmer, 2006). In avian species, the growth hormone is involved in secondary features such as egg production, aging and reproduction (Zhao *et al*., 2004). Buggioti and Primmer (2006) identified many amino acids in avian growth hormone to be under purifying selection and concluded that growth hormone is under purifying selection following the bursts of rapid changes.

5.2.2 Growth hormone receptor

In this study, growth hormone was found to be under purifying selection in all lineages (Figure 4.12). However, previous studies have found that variable evolution rates in growth hormone (GH) correlate with changes in evolution rate of growth hormone

receptor (GHR). This is probably because in the receptor, the pattern is different for the extracellular and intracellular domains. In the extracellular domain, the evolution was slow followed by bursts of change during the evolution of artiodactyls, primates and rodents (Zhenfang *et al*., 2005). The two occasions are equivalent to the evolution that occurred in GH. In the intracellular domain, the evolution rate was high but constant. Acceleration was only found in the case of rodents. Previous studies also found the lineages leading to artiodactyls, primates, and rodents to be under positive selection. (Zhenfang *et al*., 2005).

5.2.3 Insulin-like growth factor 1 receptor

Maximum Likelihood Models identified one amino acid site in IGF1R to be positively selected. This site is on the L2 domain. This domain is important in binding IGF1 to bring about growth and development (Ward *et al*., 2000).

There are only a few amino acid sites responsible for the positive selection acting on a gene (Yokoyama and Yokoyama (1996). This is because different amino acid sites are under different selective pressure because the sites have different biological functions and structural roles.

The ectodomains of insulin-like growth factor 1 receptor (IGF1R) has two homologous domains (L1 residues 1-150 and L2 residues 300-460) which are on the N-terminal (Ward *et al*., 2000). These are separated by a cysteine-rich region. This region covers two-thirds the IGF1R molecule, and makes contact with the L1 domain but a slight contact with the L2 domain. This creates a space in the middle which is sufficient to

accomodate the ligands (Ward *et al*., 2000). Insulin-like growth factors are involved in prenatal and postnatal growth and development (Delafontaine *et al*., 2004). In avian species, growth hormone binds growth hormone receptor indirectly on the liver membrane. This then activates insulin-like growth factor 1 (IGF1) secretion. When IGF1R binds IGF1, the physiological response to this ligand is initiated in vivo (LeRoith *et al*., 1995). This stimulates growth through differentiation and proliferation of the bone and muscle cells (Kuhn *et al*., 2002). The location of the positively selected site on this domain means that this receptor-ligand binding will be enhanced and so will the growth. Domestication probably played a role in the positive selection whereby growth was selected for.

Maximum likelihood methods have also been successfully used to identify amino acid sites subject to positive selection in mammalian alpha defensins. Defensins are antimicrobial peptides which are cationic in nature (Lynn *et al*., 2004). They are broad spectrum acting against fungi, enveloped viruses and both gram-positive and gram-negative bacteria. Their mode of action is by disruption of microbes' membrane integrity (Lynn *et al*., 2004). Alpha defensins are variable in their action towards different pathogens. Some microbes are sensitive to some defensins and not others. This can be caused by small changes in the primary structure of the potent defensins. These sites have been found to be under positive selection. Defensins are encoded first as prepropeptides which undergo proteolytic cleaving. The proregion is cytoprotective and is inactive. On cleaving they release the C-terminal antimicrobial peptide which is active. The sites under selection are on the mature antimicrobial peptide region.

As mammals evolved to occupy new environments, it is likely that they encountered new different pathogens. The pathogens have evolved mechanisms to escape the antimicrobial molecules. This ongoing 'arms race' with pathogens may be the significant force driving the evolution of alpha defensins (Lynn *et al*., 2004). Positive selection would be of selective advantage to the antimicrobial peptides as they would have sensitivities to new infectious agents.

As with the mammalian alpha defensins, positive selection on the IGF1R also occurred on functionally important regions.

One of the objectives of studies such as these is the expectation that there is an association between these regions under selection and phenotypes of interest (Nielsen *et al*., 2007).

In a study done by Ojeda et al., 2008, he identified a signature of section in the IGF2 gene in three pig breeds (Pietrain, Hampshire, and Duroc). These breeds are commonly used to improve growth and meat leanness in pigs.

5.3 Prediction of secondary and tertiary structures for egg production proteins

5.3.1 Prolactin

In this study, prolactin was predicted as an alpha protein with 4 long alpha helices arranged in an anti-parallel fashion (Table 4.2, Appendix 8). It also has more helices than loops. This is consistent with nuclear magnetic resonance experimental studies

done before that described the secondary structure of prolactin to be made of 50% Helices and the rest being Loops (Freeman *et al.*, 2000). Teilum *et al.*, (2005) also found the secondary structure of human prolactin to be composed of four alpha helices in his studies. It has been proposed that the 4 helices of prolactin are packed in a similar way to the porcine growth hormone. The structural similarity is based on the fact that these two genes arose from a common ancestor through gene duplication followed by divergent evolution (Forsyth and Wallis, 2002).

In this study, I found prolactin protein to have one domain which has a predicted role in receptor binding. Prolactin carries out its biological functions through binding with the prolactin receptor in a one-to-two complex (Freeman *et al.*, 2000). This is regulated by tertiary structures (Teilum *et al.*, 2005).

Prolactin binds the first subunit of the receptor with high-affinity followed by association with the second subunit which has lower affinity. This leads to an intracellular cross-phosphorylation cascade. It has been hypothesized that binding to the first subunit causes a conformational change in the hormone leading to an increased affinity for the second subunit (Freeman *et al.*, 2000). When the hormone is in the unbound state, subunit two is considered incompetent. In prolactin, helix 1, helix 4, and the second half of loop 1 contain residues important for binding site 1 (Freeman *et al.*, 2000).

The binding site is concave with residues that are aromatic and have large hydrophobic and polar residues which are positively charged (Clackson and Wells, 1995). These

complement the tryptophan and negatively charged residues of the hot spots of the receptor (Clackson and Wells, 1995).

The second receptor binding site in human prolactin is outlined by a flat hydrophobic channel which is formed between helix 1 and helix 3 (Clackson and Wells, 1995). When site 1 binds a receptor, the affinity of binding a receptor by site 2 is significantly increased approaching that of site 1 (Sivaprasad *et al*., 2004).

5.3.2 Vasoactive intestinal peptide 1 and vasoactive intestinal peptide receptor 1

From this study, vasoactive intestinal peptide 1 (VIP1) structures were found to be composed mainly of alpha helices (Table 4.3, Appendix 9). Couvineau *et al*., (2012) carried out nuclear magnetic resonance (NMR) studies on human VIP1 that also revealed that the structure has more of the alpha helices except sequence 1-5 of the N terminal which lacks a defined structure when in solution.

The biological functions of VIP1 are triggered on interaction with the vasoactive intestinal peptide receptor 1 (VIPR1) (Zhou *et al*., 2008). The N terminal 1-5 plays an important role of receptor activation mainly through the adenylyl cyclase activation.

Studies have revealed that residues in position 0, 6, 22, 24, or 28 are in contact with residues glutamine[135], aspartate[107], glycine[116], cysteine[122], and lysine[127] respectively in the N-ted of the VIPR1. During binding, the central and C-terminal residues of VIP1 forms contact with N-ted (Couvineau *et al*., (2012). The N-terminus of VIP1 is in contact with the transmembrane domains and the extracelluar loop of the receptor.

Histidine[112], leucine[131], and glutamine[134] residues present in the N-ted are among the residues involved in the VIP1-receptor binding affinity.

These three play a role in the interaction of C-terminal and the central parts, and N-ted of VIP1. The binding interaction of lysine[143], threonine[144], and threonine[147] residues of VIPR1 and histidine[1] of VIP1 is an important step in the adenylyl cyclase activation (Couvineau *et al.*, 1984).

5.4 Prediction of secondary and tertiary structures for growth proteins

5.4.1 Growth hormone and growth hormone receptor

In this study, I used Raptor X to predict the structure of growth hormone (GH). I found growth hormone to have a higher proportion of alpha helix compared to loops (Table 4.5, Appendix 11). It has a four-helix bundle which is consistent with physicochemical studies which found growth hormone to consist of a high proportion of alpha helix. This is consistent with studies done by Forsyth and Wallis, (2002) who determined the 3D structure of pig GH by X-ray to consist of a four-helix bundle. The structure of a protein is valuable in determining the functions of a protein like enzymatic activity or the interaction with other proteins (Bairoch, 2000).

GH performs many physiological actions among which is controlling growth by regulating IGF1 concentrations, stimulating lipolysis, providing FFAs, and glycerol as substrates for energy metabolism and also inhibiting insulin-induced suppression of hepatic gluconeogenesis (Clemmons, 2004). GH binds to GHR leading to growth and

metabolism (Varvio *et al*., 2008). From the computational structure prediction, GH has one domain (Somatotropin-like) which runs from residues 35-214. The function of this domain is in binding the receptor.

Helical cytokines like GH and their cognate receptors have related structures and they are likely to have a similar binding mechanism (Wells, 1996). These hormones have at least two sites for binding and oligomerizing their receptors (Wells, 1996). The GH binds and dimerizes two receptors using two different sites (Site 1 and Site 2) (Wells, 1996). GH reacts with the first receptor by using Site 1 (found at the helix IV) then binds to the second receptor using Site 2. Electrostatic and hydrophobic interactions play a significant role in the formation of the signal transduction complex (Demchuk *et al*., 1994). Although hydrophobic contacts are crucial in binding, and polar residues appear less important for affinity, polar residues are important for solubility and specificity in the binding interaction (Wells, 1996). Electrostatic properties of molecules are also important in the specificity of the hormone-receptor binding (Demchuk *et al*., 1994).

5.4.2 Insulin-like growth factor 1 and Insulin-like growth factor 1 receptor

The predicted secondary and 3D structure of IGF1R is composed of helices, beta sheets, and loops (Table 4.8, Appendix 14). The following domains were in the IGF1R structure: receptor L domain, Furin-like domain, furin-like repeats, receptor L domain, Fibronectin type 3, fibronectin type 3, fibronectin type 3, PTKc-InsR-like, and Protein tyrosine kinase.

IGF1R binds its ligand (IGF1) with high affinity consequently initiating the physiological response to the ligand (LeRoith *et al*., 1995). Comparative sequence analysis followed by confirmation through X-ray analysis has also identified the domain boundaries of the IGF1R. This has identified the N terminal half of IGF1R ectodomain to contain two homologous domains (L1 and L2) separated by a Cys-rich region (Cysteine 152-Cysteine 298) (Ward *et al*., 2000). The C terminal half of IGF1R ectodomains is made up of 3 fibronectin type III (FnIII) domains, with the second containing a large insert domain of ~120-130 residues (Ward *et al*., 2000). Intracellularly, the IGF1R has a tyrosine kinase catalytic domain (973-1229) which is flaunted by two regulatory regions: a juxtamembrane region (930-972), a C-tail with 108 residues, and residues 1230-1337 which contain the phosphotyrosine binding sites for signalling molecules (Ward *et al*., 2000).

Alanine scanning studies that were done on the IGF1R ectodomain indicated that L1 and CR domains and the carboxy-terminal peptide in the alpha subunit are components of the binding site of IGF1.Residues 131-315 in the Cys-rich plus flanking regions of L1 and L2 were found to be significant determinants of IGF1 binding. 10 residues in the L1 domain and 4 residues in the Cys-rich region which are implicated in IGF1 binding are located in two regions which are discontinuous (Ward *et al*., 2000).

The first site has amino acids asparagine11, tyrosine28, histidine30, leucine33, leucine56, phenylalanine58, arginine59, and phenylalanine90. These are distributed across the first four repeats of the L1 domain. The second functional epitope is made of Tryptophan79 which is situated in a bulge in the fourth turn of the L1 beta helix and

Arginine240, Phenylalanine241, Glutamate242, and Phenylalanine251 in the CR domain. The putative IGF1R binding site has residues which are acidic found in the corner where the Cys-rich region separates from L1. Other acidic residues of the receptor are inside the Cys-rich domain and the loop residues 255-263 which extend from module 6. Electrostatics play an important part in IGF1 binding. The C-region binds to the acidic patch of the Cys-rich region which is near L1 and the acidic patch which is in the other side of IGF1 directed to a small patch with basic residues (residues 307-310) on the L2 terminal end (Ward *et al*., 2000).

There has not been done an extensive mapping of structure-function relationships of the IGF1 and IGF1R molecules as of insulin (Meyts and Whittaker, 2002).

Mutagenesis studies of IGF1 have shown that Alanine8, Aspartate12, Phenylalanine23, and Tyrosine24 in the B-domain, Tyrosine31, Arginine36, and Arginine37 in the C-peptide and Methionine59, Tyrosine60, and Alanine62 in the A-domain are important in the high-affinity binding to the IGF1R (Meyts and Whittaker, 2002).

5.5 Predicted 3D structure comparison of prolactin, vasoactive intestinal peptide receptor 1, growth hormone, growth hormone receptor, and insulin-like growth factor 1 proteins

On doing the structure comparison of the different prolactin structures, the template modeling score (TM) value of >0.5 showed that the predicted structures of common peafowl, common pheasant, common quail, helmeted guinea fowl, Japanese quail, knob-billed duck, ostrich, wild chicken, and wild duck prolactin were highly similar. Structure

alignments with a TM value of > 0.5 are mostly in the same fold while those with TM<0.5 are not in the same fold (Xu and Zhang, 2010). Structural similarity implies that the prolactin protein may have similar functions in the different species (Xu and Zhang, 2010).

The structural alignment of Japanese quail and wild chicken vasoactive intestinal peptide receptor 1 (VIPR1) structures also showed high similarity as shown by TM value of 0.675.

The structural comparison of common quail, greylag goose, wild chicken, and wild duck growth hormone (GH) gave a TM>0.5 showing that these structures were similar. This is an indication that there has been minimal variation in the overall fold of the different species' GH. Although there has not been a direct determination of the GH structure from other species, a homology model for the sheep growth hormone suggested that there is minimal variation in the overall fold despite the significant difference in the sequences of different species (Forsyth and Wallis, 2002).

A structural alignment of greylag goose, southern ostrich, wild chicken, and wild duck growth hormone receptor (GHR) showed that these structures were highly similar as shown by TM value of 0.880.

The 3D structural comparison of Japanese quail, southern ostrich, wild chicken, and wild duck insulin-like growth factor 1 (IGF1) showed that these structures were highly similar as shown by TM value of 0.615.

Comparative analysis of protein structure is important in enabling the study of evolutionary and functional relationships between proteins (Balasubramanian *et al.*, 2012). Notably, the sequences of the different species of prolactin, VIPR1, GH, GHR, and IGF1 which had similar structures had variations in some sites of the sequences. Despite these variations, their structures were similar to each other. This is an indication that evolution conserves the structure of a protein more than the sequence (Balasubramanian *et al.*, 2012). However, there are cases where similar or identical sequences have different 3D structures (Kinch and Grishin, 2002). The 3D structure determines the biological role of a protein (Bairoch, 2000).

In addition to the different proteins having similar functions in different species, the proteins can probably bind to each others' receptors and carry out their biological functions. The similarity in structure may also mean that the proteins from the different species share a common ancestor (Kinch and Grishin, 2002).

This is probably so as is the conclusion from a study done by Balasubramanian *et al.*, (2012). In their study they compared the structures of three Superoxide dismutases from *Oryza sativa*: manganese superoxide dismutase (Mn SoD), iron superoxide dismutase (Fe SoD), and copper-zinc superoxide dismutase (Cu-Zn SoD). Mn SoD and Fe SoD were found to have similar structures with corresponding similar functions while Cu-Zn SoD had a different tertiary structure from the two. Cu-Zn SoD was also found to have evolved independently from Mn SoD and Fe SoD which were closely related as they had a common phylogenetic origin. From this study, the two enzymes with similar structures can also bind each other's ligands (Balasubramanian *et al.*, 2012).

87

5.6 Predicted 3D structure comparison of vasoactive intestinal peptide 1 and insulin-like growth factor 1 receptor proteins

The predicted 3D structures of Japanese quail, southern ostrich, wild chicken, and wild duck insulin-like growth factor 1 receptor (IGF1R) had differing structures. For vasoactive intestinal peptide 1 (VIP1), the structural comparison of the two wild chicken sequences found these to be different. On comparison of the structures, the template modeling score (TM) values were less than 0.5. This suggests that the genes in the different species have different functions.

The structures of proteins evolve through different mechanisms, often including gene duplication then mutation and selection (Kinch and Grishin, 2002). A combination of different mechanisms can bring about mutation at the structural level. Most of these mechanisms are insertion/deletion/substitution of secondary structure, circular permutation, hairpin swap, and strand invasion (Kinch and Grishin, 2002).

The sequences of the two wild chicken VIP1 structures that were compared had many indels. The indels present in the sequences analyzed may have accumulated over time resulting to the structural differences and subsequently a difference in functions (Kinch and Grishin, 2002).

It could also be possible that for the species that were compared, there is species-specificity whereby the protein cannot bind to the ligand of another specie's protein and vice versa. The variation of 3D leading to species-specificity is in support of a hypothesis from a study done by Govindaraj *et al*., (2011) on toll-like receptor 8

(TLR8). In their study, they discovered that there was specificity of ligand recognition between rodent and non-rodent TLR8 (Govindaraj *et al*., 2011). They hypothesised that structural differences between rodent and non-rodent TLR8s lead to this species-specificity. TLR8 plays a vital role in innate immune response by recognizing ssRNA derived from viruses and small molecular weight ligands respectively (Govindaraj *et al*., 2011). The mechanism of antiviral activity occurs through the binding of TLR8 through its ligand binding cavity with the ligand (Govindaraj *et al*., 2011). In their study, they discovered an undefined region LRR14-15 (residues 438-442) which varied in 3D structures. Although this region is not involved in ligand interaction, it is required for TLR8 activation. Despite both non-rodent and rodent TLR8s being able to bind their ligands, rodent TLR8 have signalling initiators which are weaker compared to those of non-rodent TLR8. Non-rodent ligands can bind with rodentTLR8 although the signalling mechanism is not activated.

Therefore, they hypothesised that variation in the 3D structure caused the species-specificity in ligand recognition as the LRR14-15 is in close proximity to the binding region (Govindaraj *et al*., 2011).

Although structural similarity may be used in elucidating function, there are cases whereby similar structures may have a divergence in function and non-similar structures have a convergence of the functions (Kinch and Grishin, 2002).

CHAPTER SIX

CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

The null hypothesis that there are no signatures of selection in candidate genes for egg production and growth in poultry was rejected by the findings in this study.

The findings supported the alternative hypothesis that there are signatures of selection in candidate genes for egg production and growth in poultry.

The general objective has also been met as I identified molecular signatures of selection at candidate genes involved in growth and egg production in poultry through the computational approach.

- **Analysis of molecular signatures of selection at production genes for eggs**

Positive selection was identified in the prolactin gene in the lineage leading to domestic birds (Figure 4.5). Human beings have attempted to increase egg production through domestication activities (Cheng, 2010). This has been a major force in shaping the DNA sequence of poultry prolactin.

On the other hand, purifying selection was identified in all lineages of vasoactive intestinal peptide receptor 1 (VIPR1) gene (Figure 4.4). When synonymous mutations occur, the sequence of the amino acid is not altered. It was long thought that these mutations had no effect on the function of the protein encoded or advantageous to the organism. However, the genes with this kind of mutation are important, highly

expressed, and are involved in key pathways (Ngandu *et al*., 2008). Therefore, VIPR1 could be an important gene in poultry egg production (Zhou *et al*., 2008).

- **Analysis of molecular signatures of selection at production genes for growth**

In growth hormone, positive selection was detected in ruminants and artiodactyls lineages (Figure 4.10). Growth is an important economic trait in Livestock production which is influenced by various genes**.** Artificial selection of growth characteristics has been practiced mainly in ruminants and artiodactyls. Hence, it has been the major driving force in shaping the growth characteristics of these species.

IGF1R gene has one amino site which is positively selected. IGF1R binds IGF1 leading to pre-natal and post-natal growth and development (Delafontaine *et al*., 2004). As the positive selected site is on the L2 domain which is important for binding IGF1, the different species have enhanced growth. This may be an important marker in breeding livestock with improved growth.

- **Prediction of secondary and tertiary structures of production genes for eggs and growth**

Protein structure prediction is important in understanding the effects of selection on the protein function (Bairoch, 2000). It is useful in designing in vitro studies for studying the effects of selection because the mechanism that determines the function is well known

.

- **Comparison of the structures of production genes for eggs and growth**

For prolactin, vasoactive intestinal peptide receptor 1 (VIPR1), growth hormone, growth hormone receptor, and insulin-like growth factor 1 (IGF1), the structures of the species compared were similar. Therefore, for each gene that showed similarity, the function of the gene is conserved across the species.

The same breeding program formulated for one species may be applied in all the other species which have similarity in structure. This will save on resources that would have been used for formulating individual breeding programs. Unfortunately for insulin-like growth factor 1 receptor (IGF1R) which had a positively selected site, the species had different 3D structures hence breeding programs for improving growth should be formulated for individual species.

6.2 Recommendations

- Experimental studies in genes under purifying selection should be carried out to determine the sites under selection and the roles they play on the function of the protein.

- Breeders should perform *in vitro* and *in vivo* studies to ascertain the effect of the amino acid selection on Insulin-like growth factor 1 receptor (IGF1R) on the phenotype. IGF1R is a potential marker for improving growth in poultry.

- *In vivo* and *in vitro* studies should be performed to determine whether genes of the different species whose predicted 3D structures are similar share the same function. Structures that are not similar could still share a function. Therefore, studies should be carried out to find out if the vasoactive intestinal peptide 1 and insulin-like growth factor 1 receptor genes with dissimilar structures could have the same function in the different species.

REFERENCES

Adams, T. E., Epa, V.C., Garrett, T.P., & Ward ,C.W. (2000). Structure and function of the type 1 insulin-like growth factor receptor. *Cell. Mol. Life Sci. 57,* 1050-1093.

Agricultural sector development strategy (2010). Retrieved from http://www.gafspfund.org/sites/gafspfund.org/files/Documents/5.%20Kenya_strategy.pdf.

Akey, J. M., Zhang, G., Zhang, K., Jin, L., & Shriver, M. D. (2002). Interrogating a high-density SNP map for signatures of natural selection. *Genome research*, *12*(12), 1805-1814.

Alipanah, M., Shojaian, K., & Bandani, H. K. (2011). The polymorphism of prolactin gene in native chicken Zabol region. *J. Anim. Vet. Adv*, *10*(5), 619-621.

Al-Nasser, A., Al-Khalaifa, H., Al-Saffar, A., Khalil, F., Albahouh, M., Ragheb, G., Al-Haddad, A. & Mashaly, M. (2007). Overview of chicken taxonomy and domestication. *World's Poultry Science Journal*, *63*(2), 285-300.

Andrade, P. C., Grossi, D. A., Paz, C. C. P., Alencar, M. M., Regitano, L. C. A., & Munari, D. P. (2008). Association of an insulin-like growth factor 1 gene microsatellite with phenotypic variation and estimated breeding values of growth traits in Canchim cattle. *Animal Genetics*, *39*(5), 480–485.

Aquadro, C. F., Bauer DuMont, V., & Reed, F. A. (2001). Genome-wide variation in the human and fruitfly: a comparison. *Current Opinion in Genetics & Development*, *11*(6), 627–634.

Bairoch, A. (2000). The ENZYME database in 2000. *Nucleic acids research*, *28*(1),

304-305.

Balasubramanian, A., Das, S., Bora, A., Sarangi, S., & Mandal, A. B. (2012).

　　Comparative Analysis of Structure and Sequences of Oryza sativa Superoxide

　　Dismutase. *American Journal of Plant Sciences*, *03*(09, 1311.

Besbes, B. (2009). Genotype evaluation and breeding of poultry for performance under

　　sub-optimal village conditions. *World's Poultry Science Journal*, *65*(02), 260–

　　271.

Bole-Feysot, C., Goffin, V., Edery, M., Binart, N., & Kelly, P. A. (1998). Prolactin

　　(PRL) and its receptor: actions, signal transduction pathways and phenotypes

　　observed in PRL receptor knockout mice. *Endocrine Reviews*, *19*(3), 225–268.

Buggiotti, L., & Primmer, C. R. (2006). Molecular evolution of the avian growth

　　hormone gene and comparison with its mammalian counterpart. *Journal of

　　Evolutionary Biology*, *19*(3), 844–854.

Chaiseha, Y., Youngren, O. M., & Halawani, M. E. E. (2004). Expression of Vasoactive

　　Intestinal Peptide Receptor Messenger RNA in the Hypothalamus and Pituitary

　　Throughout the Turkey Reproductive Cycle. *Biology of Reproduction*, *70*(3),

　　593–599.

Cheng, H.-W. (2010). Breeding of tomorrow's chickens to improve well-being. *Poultry

　　Science*, *89*(4), 805–813.

Clackson, T. & Wells, J. A. (1995). A hot spot of binding energy in a hormone–

　　receptor interface. *Science*, 267, 383–386.

Clemmons, D. R. (2004). The relative roles of growth hormone and IGF-1 in controlling

　　insulin sensitivity. *The Journal of Clinical Investigation*, *113*(1), 25–27.

Couvineau, A., Rouyer-Fessard, C., Fournier, A., St Pierre, S., Pipkorn, R., & Laburthe, M. (1984). Structural requirements for VIP interaction with specific receptors in human and rat intestinal membranes: effect of nine partial sequences. *Biochemical and biophysical research communications*, *121*(2), 493-498.

Couvineau, A., Ceraudo, E., Tan, Y.-V., Nicole, P., & Laburthe, M. (2012). The VPAC1 receptor: structure and function of a class B GPCR prototype. *Frontiers in Endocrinology*, *3*.

Cui, J. X., Du, H. L., Liang, Y., Deng, X. M., Li, N., & Zhang, X. Q. (2006). Association of polymorphisms in the promoter region of chicken prolactin with egg production. *Poultry Science*, *85*(1), 26–31.

Das, A. T., Klaver, B., & Berkhout, B. (1998). The 5′ and 3′ TAR elements of human immunodeficiency virus exert effects at several points in the virus life cycle. *Journal of virology*, *72*(11), 9217-9223.

Delafontaine, P., Song, Y. H., & Li, Y. (2004). Expression, regulation, and function of IGF-1, IGF-1R, and IGF-1 binding proteins in blood vessels. *Arteriosclerosis, thrombosis, and vascular biology*, *24*(3), 435-444.

DeLano, W. L. The PyMOL Molecular Graphics System (2002). DeLano Scientific, Palo Alto, CA, USA. Retrieved from http://www.pymol.org.

Dekkers, J. C. (2002). The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics*, *3*(1), 22-32.

Demchuk, E., Mueller, T., Oschkinat, H., Sebald, W., & Wade, R. C. (1994). Receptor binding properties of four-helix-bundle growth factors deduced from electrostatic analysis. *Protein Science : A Publication of the Protein Society*, *3*(6), 920–935.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

El-Halawani, M. E., Whiting, S. E., Silsby, J. L., Pitts, G. R., & Chaiseha, Y. (2000). Active immunization with vasoactive intestinal peptide in turkey hens. *Poultry Science*, *79*(3), 349–354.

Etherton, T. D. (2004). Somatotropic function: the somatomedin hypothesis revisited. *Journal of animal science*, *82*(13_suppl), E239-E244.

FAO, (2010).*Chicken genetic resources used in small holder production systems and opportunities for their development*. P. Sorensen, FAO smallholder poultry production paper, No. 5. Rome: FAO.

FAO, (2014). *Decision tools for family poultry development.* FAO Animal Production and Health Guidelines No. 16. Rome: FAO.

Fay, J. C., & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics*, *155*(3), 1405-1413.

Forsyth, I. A., & Wallis, M. (2002). Growth hormone and prolactin--molecular and functional evolution. *Journal of Mammary Gland Biology and Neoplasia*, *7*(3), 291–312.

Frank, S. J. (2001). Growth hormone signalling and its regulation: preventing too much of a good thing. *Growth Hormone & IGF Research*, *11*(4), 201-212

Freeman, M. E., Kanyicska, B., Lerant, A., & Nagy, G. (2000). Prolactin: Structure, Function, and Regulation of Secretion. *Physiological Reviews*, *80*(4), 1523–1631.

Gillespie JH. (1991). The Causes of Molecular Evolution. New York: Oxford Univ.

Press: 336.

Govindaraj, R. G., Manavalan, B., Basith, S., & Choi, S. (2011). Comparative Analysis of Species-Specific Ligand Recognition in Toll-Like Receptor 8 Signaling: A Hypothesis. *PLOS ONE*, *6*(9), e25118.

Hosomichi, K., Shiina, T., Suzuki, S., Tanaka, M., Shimizu, S., Iwamoto, S., Hara, H., Yoshida, Y., Kulski, J.K., Inoko, H. & Hanzawa, K. (2006). The major histocompatibility complex (Mhc) class IIB region has greater genomic structural flexibility and diversity in the quail than the chicken. *BMC genomics*, *7*(1), 322.

Houlihan PF, Goodman SM. (1986). *The Birds of Ancient Egypt*. *Aris and Phillips Ltd*. Oxford, UK.

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T. and Durbin, R. (2002). The Ensembl genome database project. *Nucleic acids research, 30*(1), 38-41

Hudson, R. R., Kreitman, M., & Aguadé, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics*, *116*(1), 153-159.

Igarashi, H., Fujimori, N., Ito, T., Nakamura, T., Oono, T., Nakamura, K., Suzuki, K., Jensen, R.T. & Takayanagi, R. (2011). Vasoactive intestinal peptide (VIP) and VIP receptors-elucidation of structure and function for therapeutic applications. *International Journal of Clinical Medicine*, *2*(04), 500.

Jiang, R. S., Xu, G. Y., Zhang, X. Q., & Yang, N. (2005). Association of polymorphisms for prolactin and prolactin receptor genes with broody traits in chickens. *Poultry Science*, *84*(6), 839–845.

Jones, D. T., Taylort, W. R., & Thornton, J. M. (1992). A new approach to protein fold

recognition. *Nature*, *358*(6381), 86-89.

Kajimoto, Y., & Rotwein, P. (1991). Structure of the chicken insulin-like growth factor I gene reveals conserved promoter elements. *Journal of Biological Chemistry*, *266*(15), 9724-9731.

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., & Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nature protocols*, *7*(8), 1511-1522.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.

Kimura M. (1983). The Neutral Theory of Molecular Evolution. New York: Cambridge Univ.Press. 367.

Kingori, A. M., Wachira, A. M., & Tuitoek, J. K. (2010). Indigenous chicken production in Kenya: a review. *International Journal of Poultry Science*, *9*(4), 309-316.

Kinch, L. N., & Grishin, N. V. (2002). Evolution of protein structures and functions. *Current Opinion in Structural Biology*, *12*(3), 400–408.

Klein, S., Morrice, D. A., Sang, H., Crittenden, L. B., & Burt, D. W. (1996). Genetic and physical mapping of the chicken IGF1 gene to chromosome 1 and conservation of synteny with other vertebrate genomes. *Journal of Heredity*, *87*(1), 10-14.

Kühn, E. R., Vleurick, L., Edery, M., Decuypere, E., & Darras, V. M. (2002). Internalization of the chicken growth hormone receptor complex and its effect on

biological functions. *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology*, *132*(1), 299–308.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. & Thompson, J.D.(2007). Clustal W and Clustal X version 2.0. *bioinformatics*, *23*(21), 2947-2948.

Leakey MD & Hay.R L (1979). Pliocene footprints in the Laetolil Beds at Laetoli, northern Tanzania. *Nature,* 278, 317-323.

Lefort, V., Desper, R., & Gascuel, O. (2015). FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Molecular Biology and Evolution*, msv150.

LeRoith, D., Werner, H., Beitner-Johnson, D., & Roberts, C. T. (1995). Molecular and cellular aspects of the insulin-like growth factor I receptor. *Endocrine Reviews*, *16*(2), 143–163.

Li, W. H., Wu, C. I., & Luo, C. C. (1985). A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Molecular biology and evolution*, *2*(2), 150-174.

Li, H. F., Zhu, W. Q., Chen, K. W., Zhang, T. J., & Song, W. T. (2009). Association of polymorphisms in the intron 1 of duck prolactin with egg performance. *Turkish Journal of Veterinary and Animal Sciences*, *33*(3), 193-197.

López, M. E., Neira, R., & Yáñez, J. M. (2015). Applications in the search for genomic

selection signatures in fish. *Frontiers in genetics*, *5*, 458.

Lynn, D. J., Lloyd, A. T., Fares, M. A., & O'Farrelly, C. (2004). Evidence of positively selected sites in mammalian alpha-defensins. *Molecular Biology and Evolution*, *21*(5), 819–827.

MacDonald, K. C. (1992). The domestic chicken (Gallus gallus) in sub-Saharan Africa: A background to its introduction and its osteological differentiation from indigenous fowls (Numidinae and Francolinus sp.). *Journal of Archaeological Science*, *19*(3), 303–318.

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. *Nature*, *351*(6328), 652-654.

Marshall F. The origins and spread of domestic animals in East Africa (2000). In: Blench RM, MacDonald KC, editors. The Origins and Development of African Livestock: *Archaeology, genetics, Linguistics and Ethnography*. London, UK: UCL press: 191–221.

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., & Šali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual review of biophysics and biomolecular structure*, *29*(1), 291-325.

Mayr, G., Domingues, F. S., & Lackner, P. (2007). Comparative analysis of protein structure alignments. *BMC Structural Biology*, *7*(1), 50.

Meyts, P. D., & Whittaker, J. (2002). Structural biology of insulin and IGF1 receptors: implications for drug design. *Nature Reviews Drug Discovery*, *1*(10), 769–783.

Moraa, G.K., Oyier, P.A., Maina, S.G., Makanda, M., Ndiema, E.K., Alakonya, A.E., Ngeiywa, K.J., Lichoti, J. & Ommeh, S.C.(2015). Assessment of phenotypic

traits relevant for adaptation to hot environments in indigenous chickens from

four agro-climatic zones of Kenya. *Livestock research for rural development*,

27(10).

Moreki JC, Dikeme R & Poroga B (2010). The role of village poultry in food security

and HIV/AIDS mitigation in Chobe District of Botswana. *Livestock Research for*

Rural Development. 22(55). Retrieved from

http://www.lrrd.org/lrrd22/3/more22055.htm.

Mwacharo, J. M., Nomura, K., Hanada, H., Han, J. L., Amano, T., & Hanotte, O. (2013).

Reconstructing the origin and dispersal patterns of village chickens across East

Africa: insights from autosomal markers. *Molecular Ecology*, *22*(10), 2683–

2697.

Naqvi, A. N. (2007). Application of molecular genetic technologies in livestock

production: potentials for developing countries. *Advances in Biological*

Research, *1*(3-4), 72-84.

National Center for Biotechnology Information (NCBI). Bethesda (MD): National

Library of Medicine (US), National Center for Biotechnology Information; (n.d.).

Retrieved from https://www.ncbi.nlm.nih.gov/genbank/.

National Center for Biotechnology Information (NCBI). Bethesda (MD): National

Library of Medicine (US), National Center for Biotechnology Information;

(n.d.).Retrieved from https://www.ncbi.nlm.nih.gov/nuccore/NM_205366.2.

National Center for Biotechnology Information (NCBI). Bethesda (MD): National

Library of Medicine (US), National Center for Biotechnology Information; (n.d.).

Retrieved from https://www.ncbi.nlm.nih.gov/gene/408184.

National Center for Biotechnology Information (NCBI). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; (n.d.).Retrieved from https://www.ncbi.nlm.nih.gov/gene/395889.

Ngandu, N. K., Scheffler, K., Moore, P., Woodman, Z., Martin, D., & Seoighe, C. (2008). Extensive purifying selection acting on synonymous sites in HIV-1 Group M sequences. *Virology Journal*, *5*(1), 160.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, *39*, 197–218.

Nielsen, R., & Yang, Z. (1998). Likelihood Models for Detecting Positively Selected Amino Acid Sites and Applications to the HIV-1 Envelope Gene. *Genetics*, *148*(3), 929–936.

Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., & Clark, A. G. (2007). Recent and ongoing selection in the human genome. *Nature Reviews Genetics*, *8*(11), 857-868.

Nyaga, P., (2007). The structural, marketing and importance of the commercial and village poultry industry: analysis of the poultry sector in Kenya. Food and Agriculture Organisation.

Ojeda A, Huang LS., Ren J, Angiolillo A., Cho IC., Soto H., Lemús-Flores C., Makuza SM., Folch JM., and Pérez-Encizo M. (2008). Selection in the making: A worldwide survey of haplotypic diversity around a causative mutation in porcine IGF2. *Genetics.* 178, 1639–1652.

Okeno, T. O., Kahi, A. K., & Peters, K. J. (2012). Characterization of indigenous chicken production systems in Kenya. *Tropical Animal Health and Production*, 44(3), 601–608.

Oleksyk, T. K., Smith, M. W., & O'Brien, S. J. (2010). Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1537), 185–205.

Olwande, P. O., Ogara, W. O., Okuthe, S. O., Muchemi, G., Okoth, E., Odindo, M. O., & Adhiambo, R. F. (2010). Assessing the productivity of indigenous chickens in an extensive management system in southern Nyanza, Kenya. *Tropical Animal Health and Production*, *42*(2), 283–288.

Peng, J., & Xu, J. (2009). Boosting protein threading accuracy. In *Research in computational molecular biology* , 31-45. Springer Berlin/Heidelberg.

Peng, J., & Xu, J. (2010). Low-homology protein threading. *Bioinformatics*, *26*(12), i294-i300.

Peng, J., & Xu, J. (2011). A multiple-template approach to protein threading. *Proteins: Structure, Function, and Bioinformatics*, *79*(6), 1930-1939

Qanbari, S., Strom, T.M., Haberer, G., Weigend, S., Gheyas, A.A., Turner, F., Burt, D.W., Preisinger, R., Gianola, D. & Simianer, H. (2012). A high resolution genome-wide scan for significant selective sweeps: an application to pooled sequence data in laying chickens. *PloS one*, *7*(11), e49525.

Rambaut, A. (2014). *FigTree v1. 4.2*. Retrieved from *http://tree. bio. ed. ac. uk/software/figtree.*

RaptorX: a Web Portal for Protein Structure and Function Prediction.

104

Retrieved from http://raptorx.uchicago.edu/.

*RaptorX Structure Alignment Server.*Retrieved from

http://raptorx.uchicago.edu/DeepAlign.

Riddle, O., Bates, R. W., & Lahr, E. L. (1935). Prolactin Induces Broodiness in Fowl. *American Journal of Physiology -- Legacy Content*, *111*(2), 352–360.

Romanov, M. N., Talbot, R. T., Wilson, P. W., & Sharp, P. J. (2002). Genetic control of incubation behavior in the domestic hen. *Poultry science*, *81*(7), 928-931.

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., … Lander, E. S. (2006). Positive Natural Selection in the Human Lineage. *Science*, *312*(5780), 1614–1620.

Selection and Biotechnology: the best of both worlds (n.d.). Retrieved from *www.ccsi.ca/Meetings/canmore/jack_dekkers.PDF*

Sivaprasad, U., Canfield, J. M. & Brooks, C. L. (2004). Mechanism for ordered receptor binding by human prolactin. *Biochemistry*, 43, 13755–13765.

Tadelle D, Million T, Alemu Y & Peters KJ (2003). Village chicken production systems in Ethiopia: 2. Use patterns and performance evaluation and chicken products and socio-economic functions of chicken; *Livestock Research for Rural Development* (15)1. Retrieved from http://www.lrrd.org/lrrd15/11tadeb151.htm

Tanaka, M., Hosokawa, Y., Watahiki, M., & Nakashima, K. (1992). Structure of the chicken growth hormone-encoding gene and its promoter region. *Gene*, *112*(2), 235-239.

Tanaka, M. I. N. O. R. U., Hayashida, Y. U. K. I. N. O. B. U., Sakaguchi, K. E. S. A. M.

I., Ohkubo, T., Wakita, M., Hoshino, S., & Nakashima, K. (1996). Growth hormone-independent expression of insulin-like growth factor I messenger ribonucleic acid in extrahepatic tissues of the chicken. *Endocrinology*, *137*(1), 30-34.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA6. Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729.

Teilum, K., Hoch, J. C., Goffin, V., Kinet, S., Martial, J. A., & Kragelund, B. B. (2005). Solution structure of human prolactin. *Journal of Molecular Biology*, *351*(4), 810–823.

Thiel, N., Zischke, J., Elbasani, E., Kay-Fedorov, P.,& Messerle, M. (2015). Viral Interference with Functions of the Cellular Receptor Tyrosine Phosphatase CD45. *Viruses*, *7*(3), 1540–1557.

Toomajian, C., Ajioka, R. S., Jorde, L. B., Kushner, J. P.,& Kreitman, M. (2003). A method for detecting recent selection in the human genome from allele age estimates. *Genetics*, *165*(1), 287-297.

USAID (2010). *Partnership for safe poultry in Kenya (PSPK) program.* Retrieved from http://pdf.usaid.gov/pdf_docs/Pnadu076.pdf.

Usman, M., Basheer, A., Akram, M., Babar, M. E., & Zahoor, I. (2014). PROLACTIN: CANDIDATE GENE FOR EGG PRODUCTION OR BROODINESS TRAITS IN CHICKEN. *Science International*, *26*(3).

Varvio, S.-L., Iso-Touru, T., Kantanen, J., Viitala, S., Tapio, I, Mäki-Tanila, A., … Vilkki, J. (2008). Molecular anatomy of the cytoplasmic domain of bovine

growth hormone receptor, a quantitative trait locus. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1642), 1525–1534.

Wang, S., Ma, J., Peng, J., & Xu, J. (2013). Protein structure alignment beyond spatial proximity. *Scientific Reports*, *3*(1448).

Ward, C. W., Garrett, T. P. J., Lou, M., Mckern, N. M., Adams, T. E., Elleman, T. C., … Chandana, E. V. (2000). The Structure of the Type 1 Insulin-Like Growth Factor Receptor. Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK6216/

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)*, *25*(9), 1189–1191.

Watterson, G. A., & Guess, H. A. (1977). Is the most frequent allele the oldest?. *Theoretical population biology*, *11*(2), 141-160.

Wells, J. A. (1996). Binding in the growth hormone receptor complex. *Proceedings of the National Academy of Sciences*, *93*(1), 1–6.

Wilkanowska, A., Mazurowski, A., Mroczkowski, S., & Kokoszyński, D. (2014). Prolactin (PRL) and prolactin receptor (PRLR) genes and their role in poultry production traits. *Folia Biologica*, *62*(1), 1–8.

Wu, S., Skolnick, J., & Zhang, Y. (2007). Ab initio modeling of small proteins by iterative TASSER simulations. *BMC biology*, *5*(1), 17.

Xu, J.,& Zhang, Y. (2010). How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics (Oxford, England)*, *26*(7), 889–895.

Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, *15*(5), 568–573.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591.

Yang, Z., Nielsen, R., Goldman, N., & Pedersen, A. M. K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, *155*(1), 431-449.

Yang, Z., Wong, W. S. W., & Nielsen, R. (2005). Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. *Molecular Biology and Evolution*, *22*(4), 1107–1118.

Yokoyama, S., & R. Yokoyama (1996). Adaptive evolution of photoreceptors and visual pigments in vertebrates. *Ann. Rev. Ecol. Syst*. 27, 543-567.

Yousefi, S., Raoufi, Z., & Rasouli, Z. (2012). Investigation of prolactin gene polymorphism in Japanese quail. *Iranian Journal of Biotechnology*, *45*.

Zhao, R., Muehlbauer, E., Decuypere, E., & Grossmann, R. (2004). Effect of genotype-nutrition interaction on growth and somatotropic gene expression in the chicken. *General and Comparative Endocrinology*, *136*(1), 2–11.

Zhenfang, H., Ying, L.,& Yaping, Z. (2005). Fast evolution of growth hormone receptor in primates and ruminants. *Chinese Science Bulletin*, *50*(10), 997.

Zheng, W. M. (2008, May). The use of a conformational alphabet for fast alignment of protein structures. In *International Symposium on Bioinformatics Research and Applications,* 331-342. Springer, Berlin, Heidelberg.

Zhou, M., Lei, M., Rao, Y., Nie, Q., Zeng, H., Xia, M., Liang, F., Zhang, D. & Zhang, X. (2008). Polymorphisms of vasoactive intestinal peptide receptor-1 gene and their genetic effects on broodiness in chickens. *Poultry science*, *87*(5), 893-903.

Zhu, Y. F., Li, H. F., Han, W., Shu, J. T., Song, W. T., Zhang, X. Y., & Chen, K. W.
(2010). Detecting Adaptive Evolution of Galliform and Anseriform Avians Mx
Genes. *Journal of Animal and Veterinary Advances*, *9*(13), 1811-1815.

Appendix 1: Homologues for prolactin gene.

Prolactin gene		
Species	**Expectation value**	**Accession Number**
1. Wild chicken 1	3e-148	BAB18728.1
2. Chicken breed 15 Hubbard	3e-148	AKQ98504.1
3. Chicken cobb 500	7e-145	AKQ98502.1
4. Chicken breed Lohmann	4e-143	AKQ98501.1
5. Chicken breed Yangshan	5e-162	AAT02223.1
6. Grey jungle fowl	1e-160	BAJ61716.1
7. Chicken breed avian 48	1e-160	AKQ98503.1
8. Wild chicken 2	5e-162	AF288765.1
9. Common quail	7e-150	BAD10927.1
10. Wild turkey	5e-147	AAB60604.1
11. White throated sparrow	3e-148	XP 005481833.1
12. Gold collared manakin	3e-148	KFW 77597.1
13. Peregrine falcon	7e-145	XP 005235773.1
14. Arctic fulmar	5e-163	KFV 94966.1
15. Greylag goose	5e-145	XP 007653890.0
16. Ostrich	4e-143	BAF81528.1
17. Wild chicken	5e-162	AAG01026.1
18. Common peafowl	1e-160	BAG68293.1

19. Japanese quail	8e-151	BAJ61717.1
20. Helmeted guinea fowl	2e-151	BAG68294.1
21. Dalmatian pelican	2e-143	KFQ 6004.1
22. Wild duck	3e-144	BAD14942.1
23. Knob billed duck	1e-143	CAJ55836.1
24. White tailed eagle 1	2e-139	KFQ 01370.1
25. Common pheasant	2e-143	BAG68292.1
26. The great cormorant	2e-143	KFW 89232.1
27. White tailed eagle 2	1e-139	XP 007867890.0
28. Rock dove	2e-134	ADK73557.1
29. Jungle crow	2e-139	BAJ61712.1
30. Ground tit	4e-139	XP 005525306.1
31. Collared flycatcher	1e-138	XP 005041658.1
32. Anna's humming bird	2e-114	XP 908765780.0
33. Zebra finch	5e-121	XP 004186110.1
34. Streamside salamander	7e-123	AP93863.1
35. Duck billed platypus	5e-114	XP 007657220.1
36. African clawed frog	6e-114	NP 001086486.1
37. Western clawed frog	5e-114	NP 001093699.1

Appendix 2: Homologues for vasoactive intestinal peptide 1 (VIP1) gene.

VIP1 gene		
Species	**Expectation value**	**Accession number**
1. Wild chicken 1	0.0	NP 990697.2
2. Swan goose	0.0	XP 008776654.0
3. Collared flycatcher	0.0	XP 005643344.4
4. Wild turkey	0.0	XP003204177.1
5. Southern ostrich 1	0.0	KFV78141.1
6. Northern fulmar	0.0	AA99877.0
7. Gold collared manakin	0.0	BAB98877
8. Rock dove	0.0	BAB99879
9. Saker falcon	0.0	AA98765
10. Peregrine falcon	0.0	KFW56765.4
11. Little egret	0.0	KFV87765.5
12. Emperor penguin	0.0	XP 004543220.0
13. Dalmatian pelican	0.0	XP 005465365.1
14. Wild chicken 2	0.0	NP 7866878.0
15. Southern ostrich2	0.0	NP 8979898.9
16. Stinkbird	0.0	NP 64763764.8
17. Medium groundfinch	0.0	KFW7878787
18. Anna's hummingbird	0.0	KFQ2334434
19. White throated sparrow2	0.0	XP 008987487.0

20. Spotted gar	0.0	XP 006574638.6
21. Ostrich1	0.0	BAB768567
22. Ostrich2	0.0	AA6758.0
23. White tailed eagle	0.0	NP 8976565.0
24. Zebra fish	0.0	BAB87765
25. Mexican tetra	0.0	BAB56432
26. Japanese ricefish	0.0	AA69870.0
27. Bicolor damselfish	0.0	XP 005644322.3
28. Olive flounder	0.0	XP 008887766.0
29. Rainbow smelt	0.0	XP 003476654.1
30. Blue damsel fish	0.0	XP 007876548.0

Appendix 3: Homologues for vasoactive intestinal peptide receptor 1 (VIPR1) gene.

VIPR1 gene		
Species	**Accession Number**	**Expectation value**
1. Wild chicken	BAA95164.1	0.0
2. Japanese quail	AED87510.1	0.0
3. Wild duck	EOA98591.1	0.0
4. Rock dove	EMC82014.1	0.0
5. Southern ostrich	BAA76574.1	0.0

6. Saker falcon1	XP 005442369.1	0.0
7. Peregrine falcon1	XP 005229590.1	0.0
8. White tailed eagle1	XP 007856847.0	0.0
9. Wild turkey	Q91085.2	0.0
10. Saker falcon2	XP 005442370.1	0.0
11. Peregrine falcon2	XP 005229591.1	0.0
12. White tailed eagle2	AAB67768.0	0.0
13. Blind cavefish	XP 007249106.1	0.0
14. Zebra fish	AAI162971	0.0
15. Nile tilapia	XP 003439239.2	0.0
16. Fairy cichlid	XP 006802769.1	0.0
17. Burton's mouthbroooder1	XP 005912343.1	0.0
18. Zebra mbuna1	XP 005463577.0	0.0
19. Zebra mbuna2	XP 006756847.8	0.0
20. Damsel fish	XP 007876487.0	0.0
21. Orange spotted spine foot	ACC78770.1	0.0
22. Japanese rice fish1	AA787879.0	0.0
23. Burton's mouthbroooder2	XP 005933737.1	0.0
24. Zebra mbuna3	XP 007878780.0	0.0
25. Japanese rice fish 2	XP 004081326.1	0.0
26. Japanese puffer1	CAC82587.1	0.0
27. Japanese puffer2	XP 003977758.1	0.0

28. Rainbow trout	AAU29499.1	0.0
29. Burton's mouth brooder3	XP 007457657.6	0.0
30. Amazon moly	XP 007548620.1	2e-159

Appendix 4: Homologues for growth hormone (GH) gene.

GH gene		
Species	**Accession Number**	**Expectation value**
1. Wild chicken	AHM95535.1	9e-35
2. Arctic fulmar	EOA56765.0	8e-67
3. Wild duck	EOA99704.1	5e-34
4. Greylag goose	AAN37412.1	9e-34
5. Common quail	ACJ73931.1	5e-34
6. Peregrine falcon	XP 005238874.1	2e-30
7. Rock dove	EMC85315.1	6e-32
8. White throated sparrow1	XP 005487217.1	2e-28
9. Medium groundfinch1	XP 005425878.1	1e-29
10. South ostrich	EOA76764.0	4e-34
11. Wild canary1	ABB56767.0	5e-33
12. Medium groundfinch2	XP 005431290.1	8e-27
13. European pied flycatcher	ABB70042.1	5e-33
14. Zebra finch1	XP 002196167.1	1e-27

15. Gold collared manakin	XP 003686878.1	5e-33
16. Wild canary2	XP 002435767.0	1e-25
17. Ground tit	XP 005524208.1	2e-29
18. Collared flycatcher	XP 005059628.1	1e-25
19. White throated sparrow2	XP 006576477.8	4e-27
20. Zebrafinch2	XP 002187284.1	3e-28
21. European rabbit	XP 007636368.7	2e-28
22. Wild Bactrian camel	XP 006177464.1	4e-23
23. Zebu	XP 001122366.7	2e-23
24. Chinese sturgeon	XP 006356356.0	2e-23
25. Russian sturgeon	ABK74674.6	9e-24
26. Sheep1	ABK59498.1	2e-23
27. Sheep2	ABO21737.1	5e-23
28. Cattle	ABK67647.0	5e-23
29. Goat	ADX66303.1	9e-24

Appendix 5: Homologues for growth hormone receptor (GHR) gene.

GHR gene		
Species	Accession Number	Expectation value
1. Chicken breed Yunnan Daweishan	KC242242.1	0

2. Chicken breed Yunnan Wuding	KC242241.1	0
3. Commercial chicken	M74057	0
4. Wild chicken1	AGG38006.1	0.0
5. Wild chicken2	NP001001293.1	0.0
6. Dalmatian pelican	XP 0876532.0	0.0
7. Greylag goose	ACY38605.1	0.0
8. Saker falcon	XP 005433804.1	0.0
9. Peregrine falcon	XP 005242027.1	0.0
10. Wild duck	ACT 20710.1	0.0
11. Knob billed duck	ACT 20711.1	0.0
12. Rock dove	EMC76968.1	0.0
13. Southern ostrich	EMC9876.0	0.0
14. Anna's humming bird	ACY2165.0	0.0
15. Golden collared manakin	ACY3476.0	0.0
16. Red throated loon	ACY5876.1	0.0
17. White throated sparrow	XP 005493766.1	0.0
18. Zebra finch	XP 002193695.2	0.0
19. Medium groundfinch	XP 005422066.1	0.0
20. Wild canary	XP 3454267876.0	0.0
21. The great cormorant	NP6473676787.0	0.0
22. Collared flycatcher	AA018173.1	0.0

Species		
23. Northern carmine bee-eater	XP 6376387878.0	0.0
24. Atlantic canary	XP 0087765409.0	0.0
25. European rabbit	1401239A	0.0
26. Wild Bactrian camel	AA987430.0	0.0
27. Alpaca	AA897322.0	0.0
28. Carolina anole1	XP 008101043.1	0.0
29. Carolina anole2	XP 0008101044.1	0.0
30. Goat	XP 0077863233.0	0.0
31. Zebu	ABM92307.2	0.0
32. Cattle	AAU94310.1	0.0
33. Sheep	NP001009323.1	0.0

Appendix 6: Homologues for insulin-like growth factor 1 (IGF1) gene.

IGF1 gene		
Species	**Accession Number**	**Expectation value**
1. Chicken haplotype h-12	JN609551.1	1e-20
2. Chicken haplotype h-11	JN609550.1	1e-20
3. Chicken haplotype h-10	JN609549.1	2e-10
4. Chicken haplotype h-9	JN609548.1	1e-40
5. Chicken breed Yunnan Daweishan	KC242240.1	3e-10

6. Chicken breed Yunnan Wuding	KC2422239.1	3e-10
7. Chicken haplotype h-8	JN593018.1	3e-25
8. Chicken haplotype h-7	JN593017.1	3e-20
9. Chicken haplotype h-6	JN593016.1	1e-40
10. Chicken haplotype h-5	JN593015.1	1e-27
11. Chicken haplotype h-4	JN593014.1	1e-35
12. Chicken haplotype h-3	JN593013.1	1e-35
13. Chicken haplotype h-2	JN593012.1	1e-35
14. Chicken haplotype h-1	JN593011.1	1e-64
15. Commercial chicken	M3279.1	4e-10
16. Wild chicken	AGG38005.1	2e-94
17. Japanese quail	AAF67202.1	2e-94
18. Great cormorant	XP 008766769.2	1e-64
19. Wild turkey	XP 003202426.1	7e-95
20. Wild duck	ABS76279.1	2e-94
21. Zebra finch	XP 006754322.0	2e-62
22. Collared flycatcher	XP 005040114.1	7e-94
23. Golden collared manakin	XP 006921111.4	1e-63
24. Medium ground finch	XP 005421104.1	1e-93
25. Dalmatian pelican	AAF98765.0	2e-94
26. Southern ostrich	AAF34222.0	3e-75

27. Wild Bactrian camel1	XP 006186100.1	1e-64
28. European rabbit1	XP 008254938.1	2e-62
29. Cattle1	AAF56222.1	2e-62
30. Goat	BAB77524.1	3e-75
31. Sheep	ACG49835.1	1e-72
32. Wild Bactrian camel2	XP 006186101.1	5e-52
33. Wild Bactrian camel3	XP 006186102.1	7e-66
34. Wild Bactrian camel4	XP 006186103.1	1e-64
35. Wild Bactrian camel5	XP 006186104.1	6e-79
36. Cattle2	AAF22156.2	1e-63
37. European rabbit2	XP 008254939.1	7e-50
38. Cattle3	AAF42111.0	7e-50
39. Cattle4	AAF11114.2	3e-75
40. European rabbit3	XP 008254940.1	2e-62
41. European rabbit4	XP 008254941.1	5e-77
42. Cattle5	AAF75333.2	3e-75
43. Cattle6	AAF73432.0	3e-75
44. Cattle7	AAF23407.1	2e-62

Appendix 7: Homologues for insulin-like growth factor 1 receptor (IGF1R) gene.

IGF1R gene		
Species	**Accession Number**	**Expectation value**
1. Wild chicken	AGG38009.1	0.0
2. Japanese quail	BAF73401.1	0.0
3. Saker falcon1	XP 005436689.1	0.0
4. Peregrine falcon1	XP 005242493.1	0.0
5. Saker falcon2	XP 005436690.1	0.0
6. Peregrine falcon2	XP 005242494.1	0.0
7. Medium ground finch1	XP 005424278.1	0.0
8. Zebra finch	XP 002199843.1	0.0
9. Medium ground finch2	XP 005424279.1	0.0
10. Golden collared manakin1	XP 004687532.2	0.0
11. Golden collared manakin2	XP 000997654.8	0.0
12. Golden collared manakin3	EMC 77848.3	0.0
13. Wild duck	EOB07472.1	0.0
14. Little egret	EOB23699.0	0.0
15. Rock dove	EMC 77329.1	0.0
16. Southern ostrich	EMC 76589.5	0.0
17. Wild turkey	XP 0032009598.1	0.0
18. Atlantic canary	XP 0034509876.0	0.0
19. White throated sparrow	XP 0012567876.5	0.0

20. Anna's humming bird	XP 0035779654.0	0.0
21. Budgerigar	XP 0045885434.3	0.0
22. Cuckoo roller	XP 0056328797.1	0.0
23. Bar tailed trogon	XP 0011187072.0	0.0
24. Cattle1	XP 0078656766.5	0.0
25. European rabbit1	XP 0066666988.2	0.0
26. European rabbit2	XP 0011765445.3	0.0
27. European rabbit3	XP 0044498885.2	0.0
28. Sheep	XP 0040085983.1	0.0
29. Cattle2	XP 0067333333.3	0.0
30. Cattle3	XP 0078899906.2	0.0

Appendix 8: Predicted protein 3D structures of prolactin.
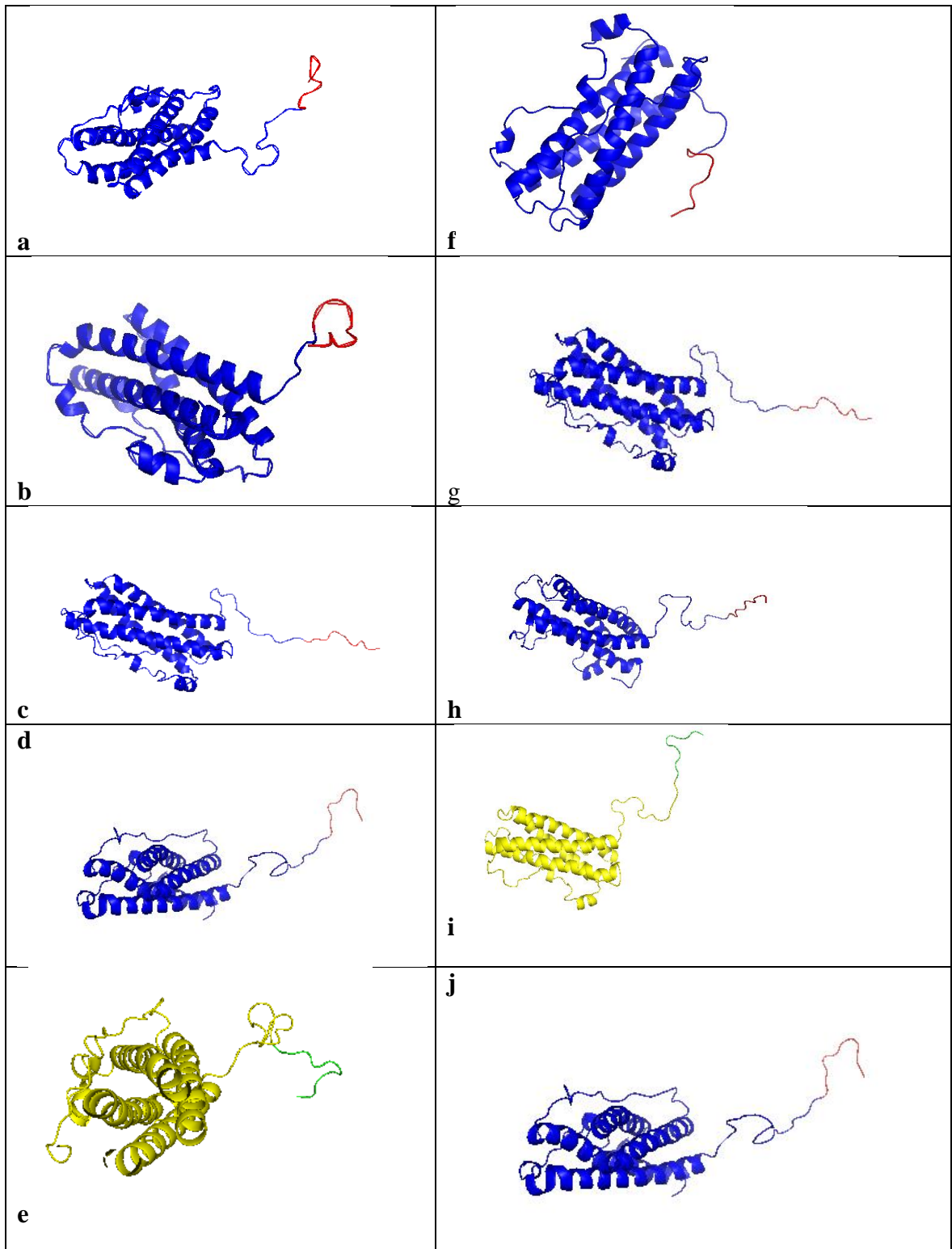
a= Predicted 3D structure of common peafowl prolactin.

b= Predicted 3D structure of common pheasant prolactin.

c= Predicted 3D structure of common quail prolactin.

d= Predicted 3D structure of helmeted guinea fowl prolactin.

e= Predicted 3D structure of wild turkey prolactin.

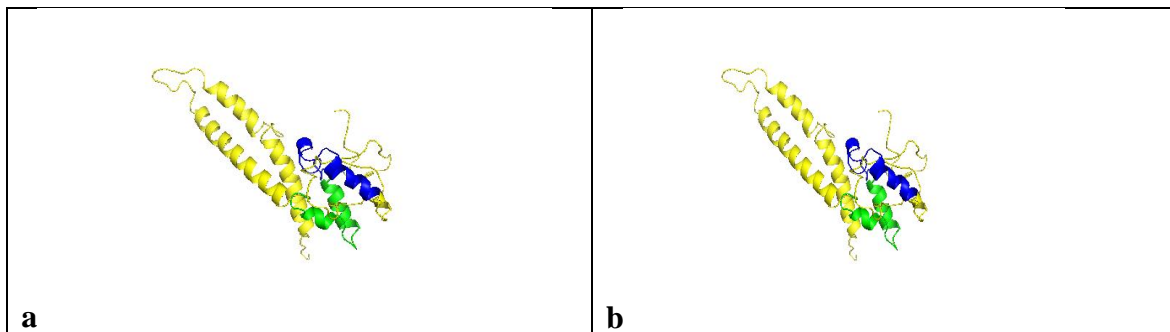**f= Predicted 3D structure of Japanese quail prolactin.**

g= Predicted 3D structure of knob-billed duck prolactin.

h= Predicted 3D structure of ostrich prolactin.

i= Predicted 3D structure of wild duck prolactin.

j= Predicted 3D structure of wild chicken prolactin.

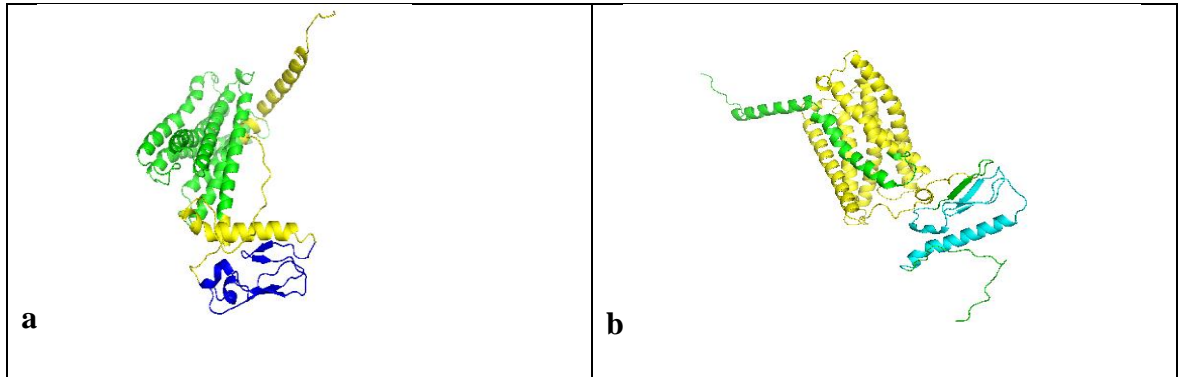**Appendix 9: Predicted protein 3D structures of vasoactive intestinal peptide 1 (VIP1).**

a   b

KEY:

a= Predicted 3D structure of wild chicken 1 VIP1.

b= Predicted 3D structure of wild chicken 2 VIP1.

Appendix 10: Predicted protein 3D structures of vasoactive intestinal peptide receptor 1 (VIPR1).
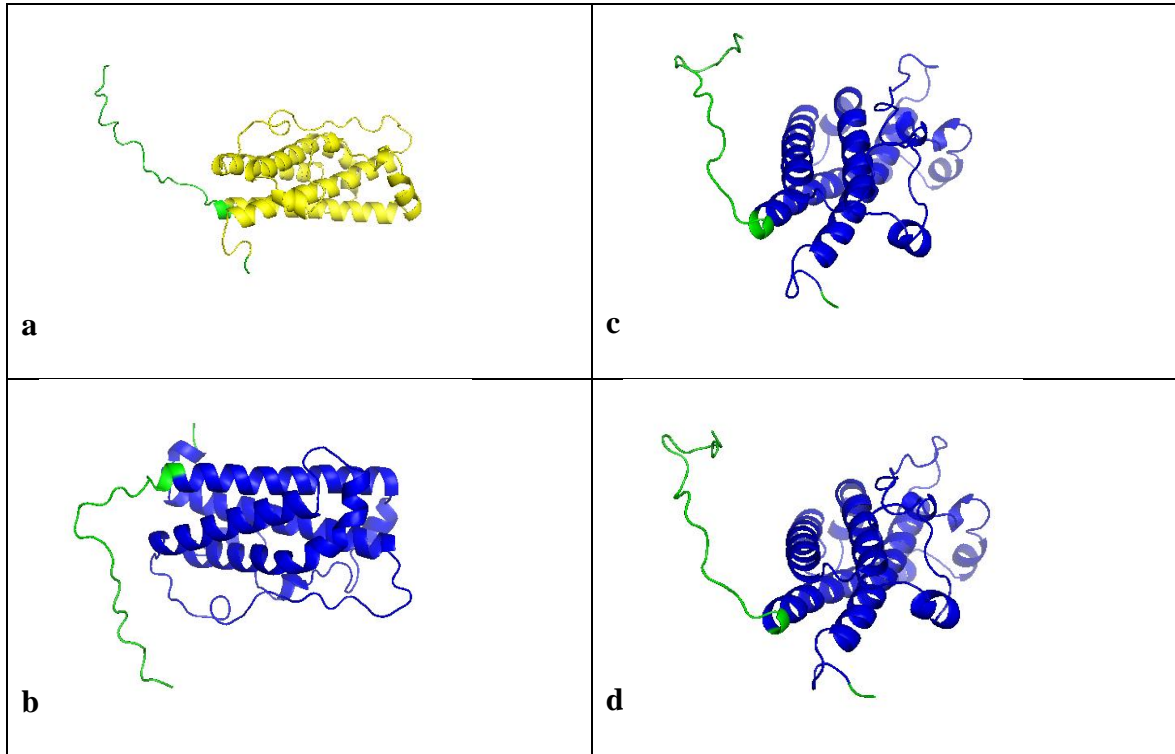

KEY:

a= Predicted 3D structure of Japanese quail VIPR1.

b= Predicted 3D structure of wild chicken VIPR1.

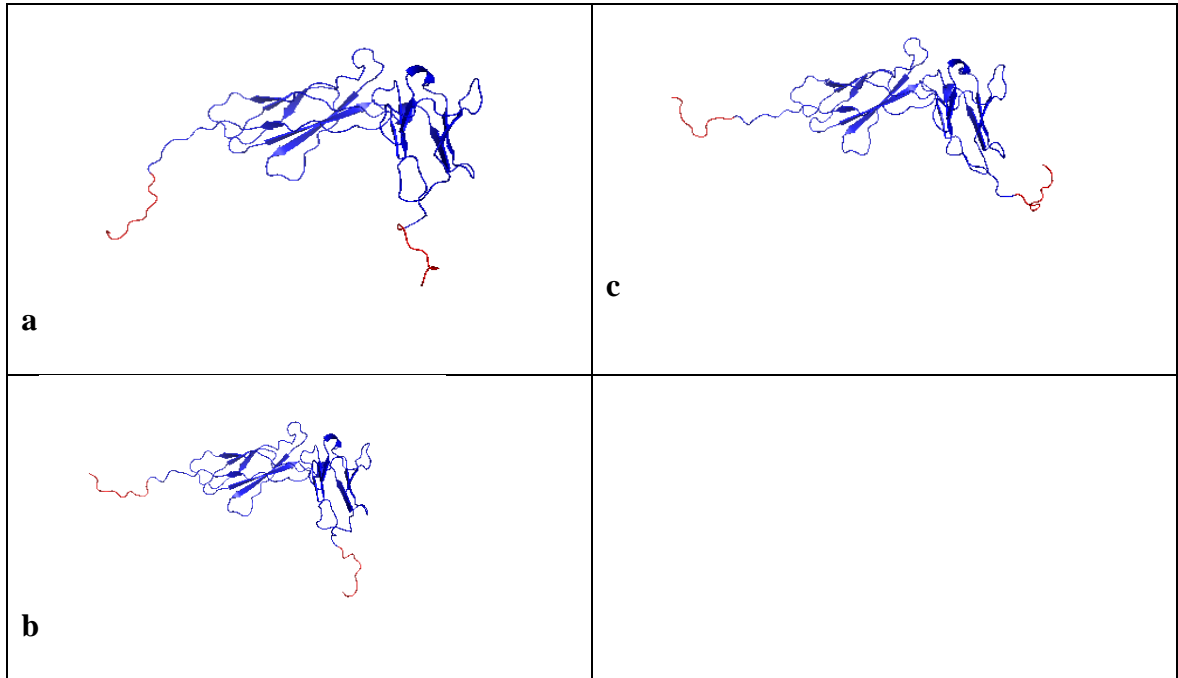**Appendix 11: Predicted protein 3D structures of growth hormone (GH).**

KEY:

a= Predicted 3D structure of common quail GH.

b= Predicted 3D structure of wild chicken GH.

c= Predicted 3D structure of greylag goose GH.

d= Predicted 3D structure of wild duck GH.

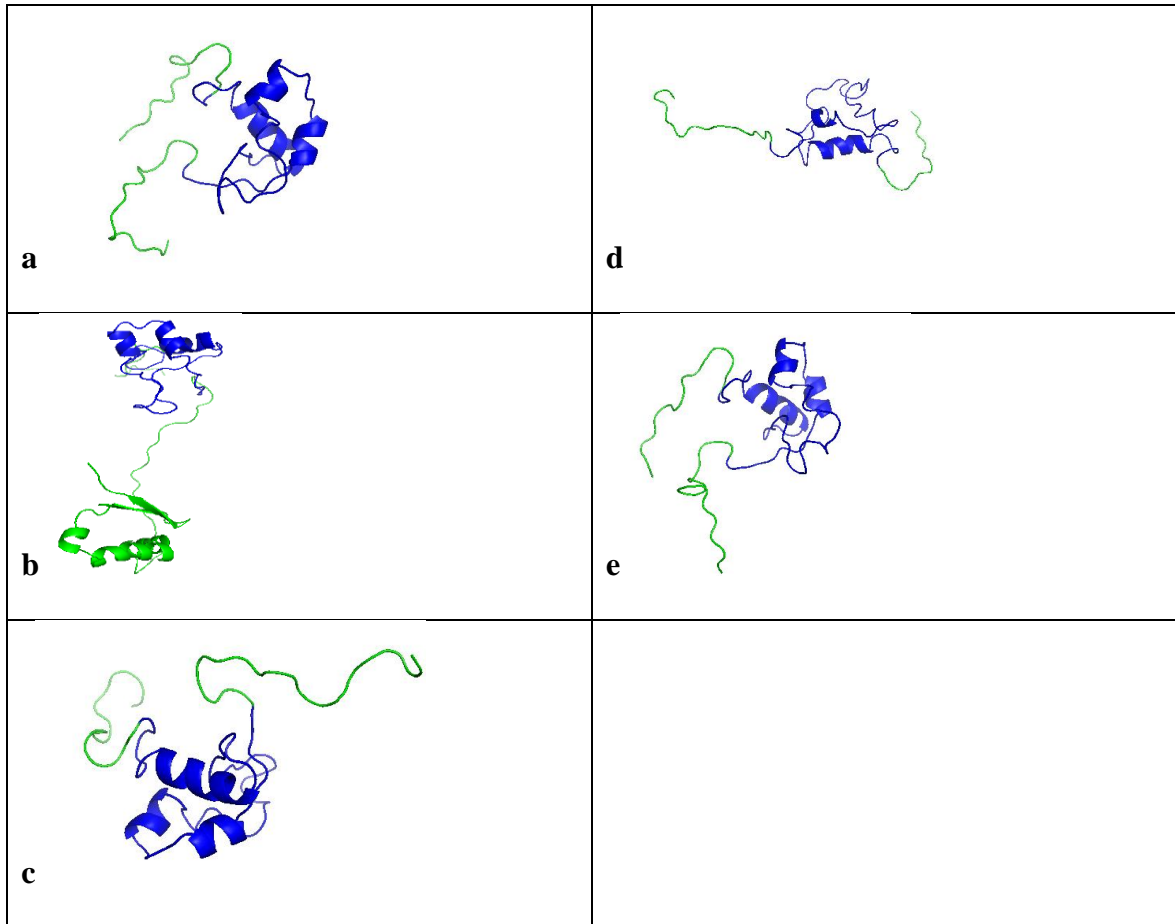**Appendix 12: Predicted 3D structure for growth hormone receptor (GHR).**

KEY:

a= Predicted 3D structure of wild chicken 1 GHR.

b= Predicted 3D structure of wild chicken 2 GHR.

c= Predicted 3D structure of wild duck GHR.

Appendix 13: Predicted protein 3D structures of insulin-like growth factor 1 (IGF1).

KEY:

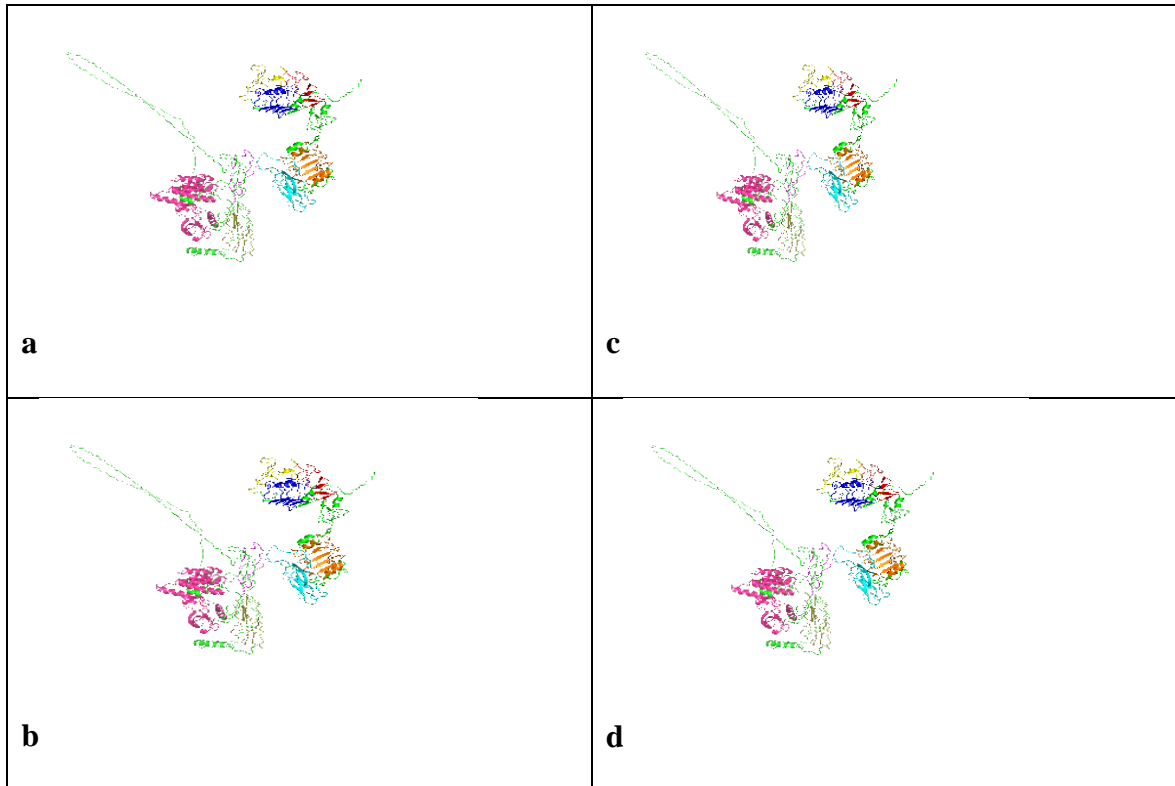**a= Predicted 3D structure of Japanese quail IGF1.**

b= Predicted 3D structure of southern ostrich IGF1.

c= Predicted 3D structure of wild chicken IGF1.

d= Predicted 3D structure of wild duck IGF1.

e= Predicted 3D structure of wild turkey IGF1.

Appendix 14: Predicted protein 3D structures of insulin-like growth factor 1 receptor (IGF1R).

KEY:

a= Predicted 3D structure of Japanese quail IGF1R.

b= Predicted 3D structure of southern ostrich IGF1R.

c= Predicted 3D structure of wild chicken IGF1R.

d= Predicted 3D structure of wild duck IGF1R.